

EMPIRICALLY BRIDGING INDIVIDUAL DIFFERENCES ACROSS
STATISTICAL LEARNING AND LANGUAGE

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Jennifer Blair Misyak

August 2012

© 2012 Jennifer Blair Misyak

EMPIRICALLY BRIDGING INDIVIDUAL DIFFERENCES ACROSS STATISTICAL LEARNING AND LANGUAGE

Jennifer Blair Misyak, Ph. D.

Cornell University 2012

Statistical learning—the process of extracting patterns from distributional properties of the input—has been proposed as a key mechanism for acquiring knowledge of the probabilistic dependencies intrinsic to linguistic structure. While such a view would predict that greater sensitivity to statistical structure should lead to better language performance, this theoretical assumption has rarely been tested empirically.

Accordingly, the work presented in this thesis is among the first to establish empirical links between statistical learning and language through the framework of studying individual differences. Contrary to assumptions that incidental learning abilities are invariant across individuals, the first small-scale individual-differences study reported systematic differences in statistical learning among normal adults, which were substantially correlated with broad cognitive measures, including language comprehension.

In two subsequent studies, a novel experimental paradigm (the AGL-SRT; Misyak, Christiansen, & Tomblin) was used to probe for within-subjects associations between individual differences in statistical learning and online sentence processing. The findings point to an overall positive relationship between individual differences in the statistical learning of adjacent or

nonadjacent dependencies and learners' processing for corresponding types of structures occurring in natural language (such as for local and long-distance dependencies entailed by subject-object relatives and subject-verb agreement sentences).

However, the complexity of the pattern of interrelations observed throughout the three studies also suggests that language and statistical learning may be related in more intricate, and sometimes counterintuitive, ways than traditionally supposed. Through discussion of theoretical implications, it is claimed that future efforts to empirically bridge together differences in statistical learning with variations in language should aid in elucidating further the broad perceptual-cognitive processes upon which statistical learning and language mechanisms may commonly supervene.

BIOGRAPHICAL SKETCH

Jennifer Blair Misyak graduated magna cum laude from Williams College in Williamstown, Massachusetts, where she received a Bachelor of Arts in Psychology and Philosophy, and was the college's first concentrator in Cognitive Science. She also studied at the University of Oxford, through an affiliation with Exeter College. Jennifer's doctoral graduate work was completed at Cornell University in Ithaca, New York, under the guidance of Morten H. Christiansen. Her research investigates individual differences in mechanisms for language and statistical learning, with broader interests relating to language evolution, normative genetic variation, computational models, and the philosophy of cognitive science.

Dedicated to the memory of Mark Testa, 1989-2012

ACKNOWLEDGMENTS

I will always be grateful to my advisor, Morten Christiansen, for his incredible mentorship and tremendous support throughout my graduate endeavors. His ideas and guidance have indelibly shaped my development as an academic scientist, and are interwoven into the fabric of my graduate experience. The underlying research and insights of this dissertation would not have been possible without him.

Special thanks are due to other members of my graduate committee, Michael H. Goldstein, David Pizarro, and Michael Spivey, for their support and example as exemplary scientists. Through collaborations, shared insights, instruction, and enthusiasms distributed across them, they have substantially strengthened my graduate experience.

I would like to warmly acknowledge Nick Chater and James Cutting; the journal manuscripts that were included in this thesis have benefitted from their feedback. My thanks also goes out to the Cornell psychology community, especially many members of the Cognitive Neuroscience Laboratory who have served as friends, collaborators, and true colleagues: Esther van den Bos, Florencia Reali, Rick Dale, Thomas Farmer, Ethan Jost, Stewart McCauley, Chris Conway, and Luca Onnis.

I also remain deeply grateful to my friends and family—especially my parents, in their unwavering support of my crazy ambitions and for their love.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgments	v
Table of Contents	vi
List of Figures	vii
List of Tables	viii
Chapter 1. Introduction	1
Chapter 2. Statistical Learning and Language: An Individual Differences Study	25
Chapter 3. When Better Statistical Learning Leads to Poorer Language Processing	74
Chapter 4. Sequential Expectations: The Role of Prediction-Based Learning in Language	101
Chapter 5. Summary and Discussion	130
Appendix	152

LIST OF FIGURES

<i>Figure 1.1.</i>	Relationship(s) between statistical learning and language	12
<i>Figure 2.1.</i>	Artificial grammars used to assess statistical learning	38
<i>Figure 2.2.</i>	Bivariate fit of adjacent and nonadjacent learning scores	49
<i>Figure 3.1.</i>	Pattern of mouse clicks for a single AGL-SRT trial	79
<i>Figure 3.2.</i>	Group learning trajectory for adjacencies	83
<i>Figure 3.3.</i>	Group reading times for subject-verb agreement sentences	89
<i>Figure 3.4.</i>	Correlation between adjacency learning and reading times	89
<i>Figure 3.5.</i>	Group reading times for subject-object relative clauses	90
<i>Figure 4.1.</i>	Grid display for presenting AGL-SRT stimulus strings	107
<i>Figure 4.2.</i>	Sequence of mouse clicks associated with a stimulus string	108
<i>Figure 4.3.</i>	Group trajectory and prediction accuracy for nonadjacencies	111
<i>Figure 4.4.</i>	Comparison of trajectories for SRNs and humans	116
<i>Figure 4.5.</i>	Prediction task scores for SRNs and humans	117
<i>Figure 4.6.</i>	Correlation between nonadjacency learning and reading times	122
<i>Figure 4.7.</i>	Group-split reading times for subject and object relatives	122

LIST OF TABLES

<i>Table 2.1.</i>	Descriptive statistics for individual-differences measures	35
<i>Table 2.2.</i>	Sentence examples for language comprehension sets	42
<i>Table 2.3.</i>	Intercorrelations between task measures	50
<i>Table 2.4.</i>	Correlations among language, statistical learning and memory	52
<i>Table 2.5.</i>	Regression coefficients of predictors for language measures	52
<i>Table 3.1.</i>	Transition probabilities for biconditional grammar	78

CHAPTER 1

Introduction

Individuals differ substantially in their language use and abilities. They differ not only throughout development and across diverse socioeconomic strata, but also among the fairly homogenous group of college-aged adults recruited for many psycholinguistic studies. Yet, despite this diversity, these variations have often been treated as peripheral phenomena or “noise,” rather than as privileged entry points into elucidating core aspects of language processes. Such relegation of individual differences bypasses powerful opportunities for theory construction and testing (Cronbach, 1957; Underwood, 1975), and for resolving “standing controversies of the text-books” (Titchener, 1910, p. 418). Conversely, understanding the sources for these linguistic variations promises to offer valuable insights into the cognitive architecture and mechanisms supporting language. Explanations at the level of the individual are also vital for forging stronger accounts of language learning and use, and the processes by which it unfolds.

A sampling of theoretical issues in the languages sciences illustrates these points more concretely. One outstanding issue among psycholinguists has been whether language processes transpire via general-purpose or linguistically specialized mechanisms, and accordingly, whether the organization of the language system is best construed as highly interdependent with or computationally autonomous (i.e., modular) from other aspects of cognition. Studying individual differences might aid in

adjudicating the matter. That is, the extent to which general cognitive and perceptuo-motor factors are profoundly implicated in language variation argues correspondingly in favor or against an interactive view. Some initial leads in this direction point towards the scenario whereby language is intimately enmeshed with basic perceptual-cognitive mechanisms, rather than being grounded solely in language-specific mechanisms (see Farmer, Misyak, & Christiansen, in press, for an individual-differences overview).

Similarly, knowledge about learner differences may also have implications for scientific accounts of whether the same or separate systems subserve the processing of linguistic dependencies. Thus, Friederici and colleagues (Friederici, 2004; Friederici, Bahlmann, Heim, Schibotz, & Anwender, 2006) argue for a functional (and localized) differentiation between humans' abilities to process finite-state transitions (adjacent dependencies) and more complex structure, such as long-distance dependencies, in natural language. Alternatively, connectionist and probabilistic learning approaches have been probing the extent to which a (singular) learning mechanism can naturally accommodate the processing of both kinds of structure via the discovery of distributional regularities (e.g., Elman, 1991; see also the discussion on *statistical learning* below). Some success of simple recurrent network simulations in this regard have also been shown to closely mirror human strengths and difficulties in language processing (e.g., Christiansen & MacDonald, 2009; MacDonald & Christiansen, 2002), thus presenting additional compelling evidence for the feasibility of a unified system. Yet, within individuals, do such skills for tracking adjacent and

nonadjacent structure cohere tightly, as differing aspects derived the same ability, or are they mostly separable?

Last but not least, investigating variations among language users might inform debates regarding the relative role of biological and experiential factors. To what extent do differences arise as the result of variations in language exposure, versus variations in the capacities of language-linked faculties? For instance, when accounting for individual differences in adults' language processing, two broad classes of explanation can be gleaned. A traditional perspective attributes these variations to constraints in cognitive resources, whereas an alternative perspective emphasizes the contributions of exposure-related factors over time. Thus, for example, when explaining individual differences in relative clause processing, proponents of the "capacity-based" view have argued for limitations imposed by language-external memory resources (Just & Carpenter, 1992; Waters & Caplan, 1996), whereas proponents of the "experience-based" view have argued for processing skills shaped by experiences with language (MacDonald & Christiansen, 2002; Wells, Christiansen, Race, Acheson, & MacDonald, 2009).

Meanwhile, expanding research into "statistical learning"—the process of extracting patterns from distributional properties of the input—provides yet another entrypoint into these discussions. Statistical learning has been proposed as a key mechanism for acquiring knowledge of the probabilistic dependencies intrinsic to linguistic structure (Gómez & Gerken, 2000; Saffran, 2001, 2003). Successful demonstrations of statistical learning abilities in the 90's (e.g., Gómez & Gerken, 1999; Saffran, Aslin, & Newport, 1996; Saffran,

Johnson, Aslin, & Newport, 1999) have spawned what McMurray and Hollich (2009, p. 365) refer to as “a small cottage industry” in which researchers use statistical learning paradigms to catalogue human sensitivity to statistics with potential relevance to a panoply of linguistic contexts. Might individual differences in statistical learning abilities themselves, then, drive differences on language tasks? While a distributional learning perspective would predict that greater sensitivity to statistical structure should lead to better language learning and performance, this decades-long theoretical assumption has rarely been tested empirically.

The role of this thesis, therefore, is to empirically bridge together individual differences in language with variations in statistical learning as a test of their interrelatedness. The next chapter introduces one of the first individual-differences studies to probe for systematic variation in statistical learning abilities. Subsequent chapters then more closely examine patterns of interrelationships between individuals’ statistical learning of dependencies and their on-line language processing. The results challenge the supposition that better statistical learning always enhances language performance, while simultaneously underscoring a tight coupling between the two processes. In doing so, this dissertation will also present findings of interest towards addressing the three theoretical issues enumerated above. That is, statistical learning is also a domain-general mechanism (albeit with potential perceptual modality subsystems; Conway & Christiansen, 2006, 2009) for learning from experience. So empirically linking these skills to outcomes in tracking both local and long-distance linguistic dependency-structures carries import for

broader language (and cognition) theories. Next, as a foundation for the chapters that follow, we begin with some background preliminaries on statistical learning phenomena and assumptions. Afterwards, an overview of the remaining chapters is provided, along with background on relevant paradigms.

Statistical learning of dependencies

Language abounds with surface-level statistical cues to structure.

Accordingly, there has been a history of theorizing within information theory and structural linguistics about the availability of such distributional statistics for identifying structural units in language (e.g., Bloomfield, 1933; Harris, 1955; Shannon, 1948). Crucially, this insight carried new significance for psychology as researchers began implementing statistical learning models demonstrating the utility of these cues for language acquisition, and also conducting experiments to establish that infants and adults could incidentally detect these statistical regularities in sound sequences. As the first behavioral demonstrations of statistical learning proper, Saffran and colleagues (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, et al., 1996; Saffran, Newport, & Aslin, 1996) showed that humans were sensitive to a type of statistical cue—the *relative* probabilities of one syllable following another syllable—in order to successfully delineate boundaries between words in a novel artificial speech stream.

Since then, the field of statistical learning has made strides in documenting humans' abilities to track these and similar probabilistic

relations during exposure to simplified artificial languages; researchers have drawn parallels to these abilities and the skills that might be required for natural language learning tasks, such as speech segmentation, phonetic categorization, phonotactic learning, and the discovery of syntactic relationships (see Chapter 2 for details). Such statistical learning has also been observed across a variety of nonlinguistic contexts, including visual scene processing, the parsing of human action sequences, visuomotor sequencing skills, and tactile learning (Baldwin, Andersson, Saffran, & Meyer, 2008; Conway & Christiansen, 2005; Fiser & Aslin, 2002a, 2002b; Hunt & Aslin, 2001). More recently, statistical learning has also been plausibly proposed as a basis for infants' early social understanding from behavioral patterns (Ruffman, Taumoepeau, & Perkins, 2012; see also Baldwin, Baird, Saylor, & Clark, 2001). And it is further likely that statistical learning skills may continue to be linked to newer domains in future research, given the ubiquity of sequential organization to human thought and action (Lashley, 1951).

In most experiments, statistical learning researchers have typically isolated humans' learning for either adjacent or nonadjacent dependencies in distributional input. Adjacent dependencies refer to relationships in which one element is predictive of another element that occurs immediately next in a temporal sequence (or possibly directly before, retrodictively, in the case of backwards transitional probabilities; e.g., Pelucchi, Hay, & Saffran, 2009; Perruchet & Desauty, 2008; see also Jones & Pashler, 2007). Nonadjacent dependencies refer to relationships in which one element is predictive of another element that occurs later in the sequence, with one or more elements

intervening between the dependency. Sensitivity to both kinds of predictive dependencies is crucial for language because key linguistic structures are conveyed through local (adjacent) and long-distance (nonadjacent) relationships. For instance, adjacent dependencies occur between syllables to form words, between words within phrases (e.g., English articles precede and predict nouns), and between phrases themselves. Analogously, nonadjacent dependencies occur across morphemic units (e.g., between an auxiliary and verb inflection) and in relating phrases across clausal embeddings. Furthermore, some frequently occurring adjacent dependencies, such as tracking the number agreement between a subject and an immediately proximal verb (e.g., *the girl*_[subject] *plays*_[verb] *with fire*), can become nonadjacent dependencies by embedding additional linguistic material (e.g., *the girl*_[subject] *with the dragon tattoo* *plays*_[verb] *with fire*).

Both forms of adjacent and nonadjacent dependency learning are believed to operate under similar computational principles and constraints from a statistical learning perspective. However, empirically, they seem to have different macro-developmental trajectories: for instance, statistical learning of nonadjacent conditional probabilities in artificial language (developing from 15 to 18 months of age; Gómez & Maye, 2005) is documented as emerging later than statistical learning for adjacent conditional (at 5.5 and 8 months-old; Johnson & Tyler, 2010; Saffran, Aslin, et al., 1996). Also, in terms of robustness at the group-level, nonadjacency learning is ostensibly more difficult than adjacency learning (e.g., Cleeremans & McClelland, 1991; Newport & Aslin, 2004)—with researchers employing a

variety of facilitative contexts to elicit successful performance, such as providing additional probabilistic cues (phonological or visual), manipulating the variability of interposed elements, scaffolding learning upon adjacent dependencies, and exploiting perceptual similarity relations (e.g., Gebhart, Newport, & Aslin, 2009; Gómez, 2002; Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007; Onnis, Christiansen, Chater, & Gómez, 2003; Van den Bos, Christiansen, & Misyak, in press). Within the implicit learning literature, it has been additionally suggested that nonadjacent dependencies may sometimes be learned more “locally,” so to speak, through the chunking of adjacent dependencies (for review, see Cleeremans, Destrebecqz, & Boyer, 1998; Perruchet & Pacton, 2006; Pothos, 2007). Yet despite these group findings, it is presently unclear how individual differences in adjacency and nonadjacency learning empirically relate to one another—a point that will be explored further in this thesis by investigating within-subjects learning for both kinds of statistical dependencies.

Statistical learning assumptions

At present, there are some generally held, nearly canonical assumptions about statistical learning. First, it has been largely considered that incidental learning of this nature is mostly invariant across people, and so we should expect to see few substantive differences from one individual to another. This follows from the central presumption, as formulated by Reber (1993), that such abilities recruit upon phylogenetically conserved and evolutionarily stable processes, and therefore should be intraindividually invariant—as well as

neurobiologically robust, common across species, and developmentally invariant (though see Misyak, Goldstein, & Christiansen, in press, for further discussion).

Secondly, a more tacit assumption widely embraced by the statistical learning community is that such learning abilities should promote natural language facility, rather than hinder performance. Indeed, despite the wide applicability of statistical learning mechanisms to other cognitive, perceptual, and motor domains, researchers' initial efforts were on testing the viability of such mechanisms as part of approaches to explain early language learning. Thus, the notion that statistical learning might be inversely related to natural language performance is exceedingly counterintuitive.

Consistent with the above assumptions, there has therefore been relatively little work in exploring whether variation in statistical learning abilities may be both systematic and informative. There are a few exceptions, with some studies examining such skills within atypical populations. A small cluster of related studies in the implicit learning literature have also probed for dissociations to psychometric intelligence, because of the presumption that any variation should be relatively little and unrelated to explicit cognition (following Reber, 1993). Among the subset of resultant findings that are least equivocal, poorer statistical learning on artificial grammar or sequence learning tasks have been evidenced among populations with agrammatic aphasia (Christiansen, Kelly, Shillcock, & Greenfield, 2010), specific language impairment (e.g., Evans, Saffran, & Robe-Torres, 2009; Tomblin, Mainela-Arnold, & Zhang, 2007), and language/learning disabilities (Grunow,

Spaulding, Gómez, & Plante, 2006; Plante, Gómez, & Gerken, 2002). Thus, the literature provides some initial indication that statistical learning abilities may be related to language outcomes, but none of these studies link beyond atypical group differences to individual differences in language processing within the typical population. While statistical learning research has flourished into the twenty-first century, there remains a conspicuous gap in empirically connecting these abilities to real-world language and cognitive outcomes (see also Romberg & Saffran, 2010). One of the most powerful ways for demonstrating empirical links—via individual differences—has largely eluded the field, because of the tacit assumption that meaningful differences in statistical learning within the normal population do not exist.

Overview of remaining chapters and paradigms

The preliminary, small-scale study of individual differences described in Chapter 2 (Misyak & Christiansen, 2012a) is among the first to explore variations in statistical learning as a principal aim.¹ Contrary to assumptions that incidental learning abilities, such as statistical learning, are invariant across individuals, we found that systematic differences among normal college-aged adults *do* exist for the statistical learning of artificial dependencies. These differences were substantially correlated with broad cognitive measures, including language comprehension. Additionally, when the effects of various factors (including fluid intelligence and cognitive

¹ Despite the recent publication date, this work was conducted before the studies in Chapters 3 and 4, and was first reported at the Annual Cognitive Science Society meeting in 2007.

motivation) were controlled for, performance on the statistical learning tasks was the only reliable predictor for accuracy in comprehending natural language sentences.

This individual-differences study was also the first to document within-subjects performances for statistical learning of both adjacent and nonadjacent dependencies (though see, Vuong, Meyer, & Christiansen, 2011, for a subsequent experiment). However, no significant correlation between adjacent statistical learning and nonadjacent statistical learning was detected.

Furthermore, there was a partly contrastive pattern of intercorrelations between each form of statistical learning (adjacency or nonadjacency) and the other study measures, including for different sets of natural language sentences. These findings raise questions for the standard conceptualization of statistical learning as a unitary system, integrating over both adjacent and nonadjacent statistics and thus theoretically sensitive to both local and long-distance dependencies in natural language. Is this undifferentiated view correct, as depicted in Figure 1.1 (*left*)? Or given the lack of reported association, might statistical learning need to be reconsidered with respect to distinct adjacent and nonadjacent manifestations, as depicted in Figure 1.1 (*right*)? Towards a better understanding in this regard, the other studies in this thesis investigate further the nature of potential links in Figure 1.1 (*right*): between adjacent statistical learning and the processing of local (and long-distant) dependencies in natural language (Chapter 3); and between nonadjacent statistical learning and the processing of long-distant dependencies in natural language (Chapter 4).

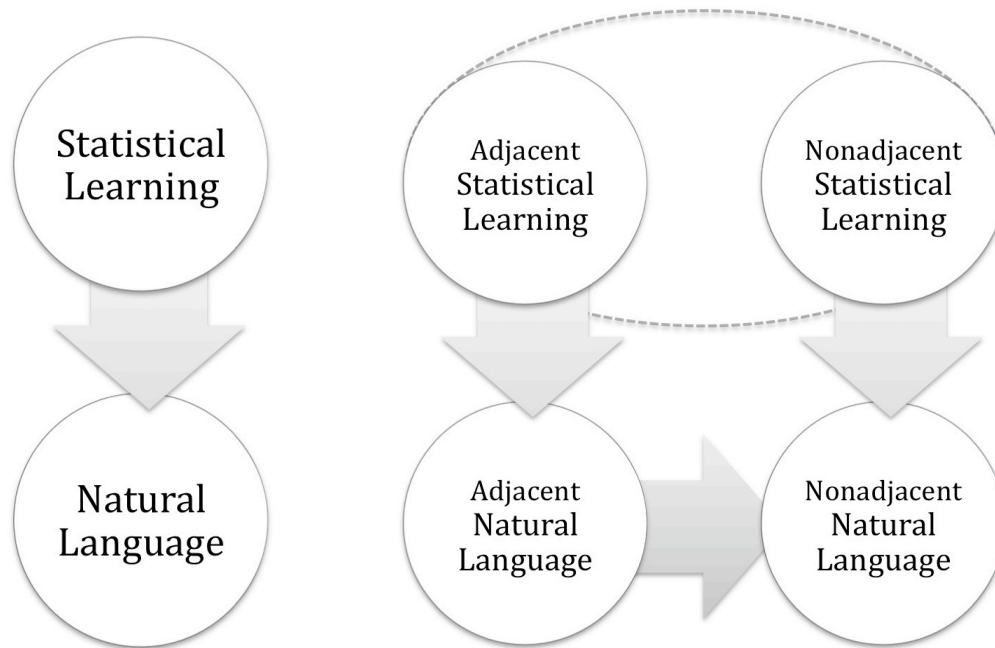


Figure 1.1. A rudimentary schematic depicting two contrasting conceptualizations of statistical learning's relationship to natural language. On the left, adjacent and nonadjacent dependency learning and processing are undifferentiated, and statistical learning influences natural language performance. On the right, the statistical learning of adjacent and nonadjacent dependencies are recognized as distinct, with their own respective influences on the processing of either adjacent or nonadjacent dependencies in natural language. A link between the two forms of natural language processing (*right-hand side*) is included to represent the relationship between processing adjacent (*viz.*, local) and nonadjacent (*viz.*, long-distant) linguistic structure. The enclosing dash-lined circle reflects the possibility that both forms of statistical learning might be subsumed within the same underlying system (see Chapter 5 for discussion).

N.B. These simplified schematics are not intended as formal models with fully specified links and relationships. The specification of main directionality is technically provisional, but arguably more consistent with results obtained from the experiments in this thesis (see, e.g., pp. 92-93 in Chapter 3) and with the theorized relationship between statistical learning and language [in the literature] that motivated these studies.

Thus, the subsequent studies in Chapters 3 and 4 investigate more closely each type of statistical learning (adjacency, nonadjacency) and its relationship to online language processing. For this purpose, a novel experimental paradigm was developed for assessing statistical learning as it unfolds: the AGL-SRT (first introduced in Misyak, Christiansen, & Tomblin, 2010). In natural rapprochement with implicit learning and sequence learning literatures (see Perruchet & Pacton, 2006), this paradigm builds upon the predominant methodologies of both traditions by incorporating design elements of artificial grammar learning (AGL; Reber, 1967) and serial reaction time (SRT; Nissen & Bullemer, 1987) tasks.

In a typical AGL task, participants observe exemplars (usually visual letter-strings; e.g., SVPTM) that, unbeknownst to them, were generated from an artificial finite-state grammar. In a subsequent test phase, participants are asked to discriminate between “grammatical” and “ungrammatical” strings on the basis of their knowledge from training. Participants often achieve above-chance classification accuracy, even when test items comprise grammatical exemplars never directly encountered in training (i.e., requiring generalizations of the grammar to new strings) and despite being unable to provide verbal reports of actual patterns or rules. Thus, participants in the AGL paradigm may evince knowledge for complex, statistical relationships, even in the absence of reported awareness for any underlying structure.

In a prototypical instantiation of the SRT task, participants respond as quickly and accurately as possible to trials of presented “targets” (e.g., illuminated lights) occurring at discrete locations on a computer screen, with

each location mapping onto a particular response button. Unknown to participants, target appearances follow a repeating (or, in some studies, probabilistic) sequence of locations. Participants become increasingly adept in anticipating and responding swiftly to targets, but their performances falter when random or incorrect targets disrupt the sequence. When target locations conform again to the training pattern, participants' performances dramatically "rebound." Because of indirect instructions and task demands for speeded responses (which discourage explicit strategizing), the SRT paradigm yields convincing demonstrations for individuals' incidental encoding of sequential relationships.

The design of the AGL-SRT (as further detailed in Chapters 3 and 4) combines methodological advantages from each of these paradigmatic mainstays in the implicit and sequence learning literatures. In essence, it does so by implementing the structured, probabilistic input of artificial grammars within a modified SRT format (without set mappings between stimuli locations and response keys). The "cover task" approach of SRT instructions discourages explicit strategizing, while engaging participants beyond the passive training component typically entailed in standard AGL experiments. The setup accommodates the use of auditory-visual strings as input and provides for the study of online trajectories of learning as it unfolds. Lastly, this methodological merger also results in sensitive indices of statistical learner differences that can be related to variations in online language processing patterns.

Accordingly, to study further the empirical relationship between statistical learning and language, we used the AGL-SRT paradigm to probe for within-subjects associations between individual differences in statistical learning and online sentence processing. One of these studies focuses on individuals' learning and processing of adjacent statistical dependencies, and is described in Chapter 3 (as a manuscript to be submitted for publication: Misyak & Christiansen, 2012b). In that study, we found that proficient *adjacent* statistical learners did not differ from others in processing long-distance natural language dependencies per se. However, greater sensitivity to adjacent statistical structure was associated with greater sensitivity to local (viz., adjacent) relations in natural language. This increased sensitivity, in turn, led to less efficient processing of long-distance dependencies (even structurally simpler ones) in certain sentence contexts. That is, when processing a sentence such as “The key to the cabinet(s) was rusty...”, better adjacent statistical learners were slower to track the nonlocal (i.e., nonadjacent) dependency (between “key... was”) when the sentence involved an intervening word (*cabinets* vs. *cabinet*) that induced a local (i.e., adjacent) dependency mismatch (between *cabinets* and *was*).

In another study (Misyak et al., 2010) presented in Chapter 4, differences in *nonadjacent* statistical learning were positively correlated with more adept processing of the long-distance dependencies contained in object-relative clause sentences (for which, in the aforementioned study, no association was found to adjacent statistical learning). That is, when processing a sentence such as “The senator that the reporter attacked admitted the

error”, better nonadjacent statistical learners were quicker to integrate the nonlocal dependency between the head-noun “*senator*” and the main verb “*admitted*” across an embedded clause (*that the reporter attacked*), whose verb also enters into a nonlocal dependency with its object “*senator*”. We also used recurrent neural networks to closely model human nonadjacency performance on the AGL-SRT task, supporting further an association-based statistical account for individual differences in long-distance dependency processing.

Together, these studies indicate that individual differences in statistical learning *are* in fact positively related to variations in language processing. At the same time, the complexity of the observed pattern of interrelations involving adjacency and nonadjacency is more intricate than traditionally supposed. As discussed further in Chapter 5, interpreting these interrelationships should prove fruitful for deepening our understanding of language and statistical learning mechanisms.

REFERENCES

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321-324.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382-1407.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, 72, 708-717.
- Bloomfield, L. (1933). *Language*. Chicago: University of Chicago Press.
- Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, 116, 382-393.
- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, 59 (Suppl. 1), 126-161.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2, 406-416.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235-253.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24-39.

- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, 17, 905-912.
- Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, 21, 561-580.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-225.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321-335.
- Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (in press). Individual differences in sentence processing. In M. Spivey, M. Joannisse & K. McRae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge, UK: Cambridge University Press.
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458-467.
- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.

- Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, 8, 245-247.
- Friederici, A. D., Bahlmann, J., Heim, S., Schibotz, R. I., & Anwender, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, 103, 2458-2463.
- Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16, 486-490.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109-135.
- Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178-186.
- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183-206.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39, 158-170.
- Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31, 190-222.

- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658-680.
- Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13, 339-345.
- Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review*, 14, 295-300.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19, 1247-1252.
- Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31, 481-507.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-131). New York: Wiley.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35-54.
- McMurray, B., & Hollich, G. (2009). Core computational principles of language acquisition: Can statistical learning do the job? Introduction to Special Section. *Developmental Science*, 12, 365-368.

- Misyak, J. B., & Christiansen, M. H. (2012a). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302-331.
- Misyak, J. B., & Christiansen, M. H. (2012b). *When better statistical learning leads to poorer language processing*. Unpublished manuscript, Cornell University, Ithaca, NY.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138-153.
- Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (in press). Statistical-sequential learning in development. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. L. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society* (pp. 886-891). Boston, MA: Cognitive Science Society.

- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113, 244-247.
- Perruchet, P., & Desauty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, 36, 1299-1305.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233-238.
- Plante, E., Gómez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35, 453-462.
- Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, 133, 227-244.
- Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 906-914.
- Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, 30, 87-104.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.

- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110-114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379-423, 623-656.
- Titchener, E. B. (1910). The past decade in experimental psychology. *The American Journal of Psychology*, 21, 404-421.
- Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, 3, 269-293.
- Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, 30, 128-134.
- Van den Bos, E., Christiansen, M. H., & Misyak, J. B. (in press). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*.
- Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and nonadjacent dependencies in statistical learning. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of*

the 33rd Annual Cognitive Science Society (pp. 964-969). Austin, TX:
Cognitive Science Society.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working
memory capacity and its relation to reading comprehension. *Quarterly
Journal of Experimental Psychology*, 49, 51-79.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M.
C. (2009). Experience and sentence processing: Statistical learning and
relative clause comprehension. *Cognitive Psychology*, 58, 250-271.

CHAPTER 2

Statistical Learning and Language: An Individual Differences Study ²

Statistical learning has been proposed as centrally connected to language acquisition and development. Succinctly defined as the discovery of structure by way of statistical properties of the input, such learning has been theorized to be robust and automatic, and observed demonstrated across a variety of both linguistic and non-linguistic contexts, including speech segmentation (Saffran, Aslin, & Newport, 1996), learning the orthographic and morphological regularities of written words (Pacton, Fayol, & Perruchet, 2005; Pacton, Perruchet, Fayol, & Cleeremans, 2001), learning artificial phonotactic patterns (Dell, Reed, Adams, & Meyer, 2000; Warker & Dell, 2006; Warker, Dell, Whalen, & Gereg, 2008), forming phonetic categories (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002), forming syntactic categories (Gerken, Wilson, & Lewis, 2005; Gómez & Lakusta, 2004), segmenting human action sequences (Baldwin, Andersson, Saffran, & Meyer, 2008), visual processing (Fiser & Aslin, 2002a, 2002b), visuomotor learning (Hunt & Aslin, 2001), tactile sequence learning (Conway & Christiansen, 2005), and non-linguistic, auditory processing (Saffran, Johnson, Aslin, & Newport, 1999; Tillmann & McAdams, 2004). But important issues still surround the general

² Misyak, J. B. & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302-331. Copyright © 2011 by Language Learning Research Club, University of Michigan. Adapted with permission.

As this chapter was adapted from the journal-approved manuscript, there may be minor deviations from the final copy-edited version; the published journal version should be consulted for verbatim quotes.

scope of statistical learning, especially with respect to how much of complex language structure can be captured by this type of learning.

Statistical learning research has sometimes also been studied as “artificial grammar learning” (AGL; Reber, 1967) or more broadly under the rubric of “implicit learning” (see Perruchet & Pacton, 2006). Such work has shown that infant and adult learners—upon brief and passive exposure to strings generated by an artificial grammar or continuous sequences of nonwords from an artificial lexicon—can incidentally acquire and evince knowledge for the predictive relations embedded within the stimuli (for reviews, see Gómez & Gerken, 2000; Saffran, 2003). Further, stimuli used within this paradigm may be devised so as to model structural properties specific to natural language, instantiating dependencies that may be characterized as either “adjacent” or “nonadjacent.” For example, Saffran (2001) documented adults’ and children’s successes in incidentally learning a simplified artificial grammar that employed predictive dependencies among *adjacent* form classes (e.g., *D-E* in the string *ADE*, where each letter represents a form class defined by a set of elements). Such relationships are characteristic of natural language, in which phrasal units may be statistically signaled by dependencies between lexical members (e.g., that determiners in English predict upcoming nouns). Similarly, Gómez (2002) investigated adults’ and infants’ learning for an artificial grammar that generated three-element strings in which initial and final items formed a *nonadjacent* dependency pair (e.g., *a-d* of *aXd*). Informed by the observation that certain elements in natural language belong to relatively small sets (function morphemes like ‘*a*,’ ‘*was*,’ ‘*-s*,’ and ‘-

ing'), whereas others belong to very large sets (open-class items such as nouns and verbs), Gómez manipulated the set size (i.e., 2, 6, 12, or 24 elements) from which she drew the middle items (Xs), and found that participants were better able to detect the nonadjacent dependencies when the variability of the middle items was at its highest (i.e., set size 24).

Given these experimental paradigms, statistical learning appears to take place using fundamentally similar computational principles and constraints within different kinds of artificial language learning (phonological, lexical, and syntactic), across concurrent levels (e.g., the simultaneous statistical learning of lexical units and syntactic phrase structure; Saffran & Wilson, 2003), and between levels (e.g., in facilitating the mapping of subsequent lexical meanings to nonwords from a statistically-segmented acoustic stream; Graf Estes, Evans, Alibali, & Saffran, 2007; Mirman, Magnuson, Graf Estes, & Dixon, 2008). Such evidence suggests that statistical learning mechanisms subserving the discovery of syntactic structure need not be distinct from those subserving the learning of nonsyntactic aspects of language such as phonology, lexicon, and semantics. However, some empirical findings have pointed to a potential distinction between forms of statistical learning that involve sequentially *adjacent* versus *nonadjacent* dependencies. Specifically, learning for these two types of dependencies have been shown to differ in their macro-level developmental trajectories and facilitative learning contexts. Within the statistical learning literature, sensitivity to nonadjacent conditional probabilities is documented later in human infancy than the earliest behavioral demonstrations of sensitivity to

adjacent conditional probabilities (see Gómez & Maye, 2005, contra Saffran et al., 1996). Additionally, compared to tracking adjacent relations, most human learners generally have a harder time tracking nonadjacent dependencies (e.g., Cleeremans & McClelland, 1991; Newport & Aslin, 2004) and require more facilitative contexts to do so successfully, such as conditions that manipulate the variability of interposed items and/or exploit perceptual similarity cues (e.g., Gebhart, Newport, & Aslin, 2009; Gómez, 2002; Onnis, Christiansen, Chater, & Gómez, 2003).

This contrast between adjacent/nonadjacent statistical learning can also be seen in how researchers have typically designed studies that isolate learning for either adjacent or nonadjacent dependencies. Accordingly, the instantiation of statistical regularities among adjacent or nonadjacent stimulus tokens in these artificial grammar tasks often aims to mirror respectively the kinds of *local* or *long-distance* relations among phonemic, lexical, and phrasal constituents that individuals process in natural language. Skill in discerning both types of artificial dependencies would therefore appear relevant for many aspects of language learning, such as segmenting words and identifying phrasal boundaries (adjacent relationships) and properly inflecting morphemes and processing embeddings (nonadjacent relationships). Yet, it is unknown if these two manifestations of statistical learning are separable skills within-individuals, rather than denoting differing aspects of the same ability. More generally, it also remains to be fully evidenced whether and to what extent statistical learning and natural language are subserved by the same underlying mechanism(s).

The present experiment therefore employs an individual-differences framework to explore the hypothesis that statistical learning and language are integrally interrelated. The aim is to document the nature of empirical interrelationships among learner differences, informed by the observation that individual differences are substantive and ubiquitous across language. To the extent that statistical learning and language are subserved by the same underlying mechanism(s), differences in language should systematically relate to and be informative of differences in statistical learning.

Next, we briefly review findings relevant to differences in statistical learning, and then discuss the individual-difference factors of specific interest in this study.

Individual Differences in Statistical Learning

To date, findings across the statistical learning and language literatures suggest that the probabilistic knowledge resulting from statistical, implicit learning processes may substantially underpin learners' acquisition of language (e.g., for a review concerning L1 development, see Gómez, 2007; for a review that relates such effects to L2 acquisition, see N. Ellis, 2002). While individual differences in language (both L1 and L2 learning/processing) have received some attention to date (for some overviews, see Bates, Dale, & Tal, 1995; Dörnyei, 2005; R. Ellis, 2004; Farmer, Misyak, & Christiansen, in press; MacDonald & Christiansen, 2002; Michael & Gollan, 2005; Vasilyeva, Waterfall, & Huttenlocher, 2008), less is known about individual differences in statistical learning within the normal population. Most evidence suggesting

the presence of systematic variation in statistical learning pertains to developmental differences, atypical populations, or from studies using putative dissociations in performance between implicit and explicit learning tasks to investigate Reber's predictions (e.g., see Reber, 1993) for implicit learning as IQ-independent and age-invariant.

Thus, although seemingly present throughout development, Saffran (2001) observed consistent performance dissimilarities between children and adults in one of her artificial language studies. Additionally, Arciuli and Simpson (2011) have reported improvements in statistical learning performance as a function of increasing age in years (from five to twelve) within typically developing children. Further, within atypical populations, performance differences on AGL or statistical learning tasks have been documented for individuals with language-related impairments: agrammatic aphasia (Christiansen, Kelly, Shillcock, & Greenfield, 2010), developmental dyslexia (Pothos & Kirk, 2004; though see counterclaims by Kelly, Griffiths, & Frith, 2002), specific language impairment (Evans, Saffran, & Robe-Torres, 2009; Hsu, Tomblin, & Christiansen, 2009), language/learning disabled adults (Grunow, Spaulding, Gómez, & Plante, 2006; Plante, Gómez, & Gerken, 2002), and Williams syndrome children and adults (albeit not after factoring group-differences in working memory or nonverbal intelligence; Don et al., 2003).

Lastly, within the normal population, some differences in AGL have been explored in relationship to psychometric intelligence. Accordingly, Reber, Walkenfeld, and Hernstadt (1991) claimed that AGL was unrelated to intelligence, as they did not detect a significant association within their study

between AGL and IQ scores from the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), nor did McGeorge, Crawford, & Kelly (1997). However, Robinson (2005) reported a negative association between WAIS-R IQ and AGL scores in a group of experienced second language learners. Conversely, other studies (Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2008; Kempe, Brooks, & Kharkhurin, 2010) showed that Culture Fair Intelligence Test (CFIT; Cattell, 1971) scores mediated successful learning on miniature second-language learning tasks bearing resemblance in their design and learning demands to those invoked by traditional AGL tasks.

Therefore, although these few studies have looked at individual differences in statistical learning (sometimes with equivocal outcomes), they have not directly sought to link such differences to variations in language abilities within the normal adult population. Finding correlations between individual differences in statistical learning and language is crucial to determining whether the two may overlap in terms of their underlying mechanisms. We thus set out to explore such associations in a comprehensive study of statistical learning and language differences using a within-subject design.

Overview of Study Measures

To determine the potential role of different types of statistical learning, we used two standard artificial grammars to isolate the learning of adjacent and nonadjacent information within-individuals. We then studied differences on these tasks in relation to differences in comprehending sentences whose

primary manipulation entails the tracking of adjacent and/or nonadjacent natural language dependencies. As the statistical learning of adjacencies and the processing of local language dependencies both require sensitivity to adjacent sequential information, we expected that measures tapping into both of these should be more strongly interrelated than potential associations obtaining between adjacent statistical learning and the comprehension of long-distance natural language structures—and analogously, similar expectations hold for the sensitivity to nonadjacent sequential information entailed by the statistical learning of nonadjacencies and the processing of long-distance language dependencies. Thus, the inclusion of both aspects of statistical learning allowed us to probe for any differential associations with our language measures, under the assumption that sensitivity to such dependencies is an integral component of language comprehension.

We also included in our study other potential contributing factors to variation across language and statistical learning. These measures were intended to assess differences in memory-related factors (verbal working memory and short-term memory), broad language-relevant variables (lexical knowledge and print exposure), and nonverbal abilities/aptitudes (fluid intelligence and cognitive motivation). Memory-related factors have become arguably the most studied individual-differences cognitive factors in the language processing literature, and so we included them here. Verbal working memory (vWM) in particular, as conventionally gauged by reading span tasks, has been correlated with native language comprehension abilities across various experiments (for a review, see MacDonald & Christiansen, 2002). It

has also begun to be extensively researched in the second language learning literature, with studies supporting an association between L2 reading span and L2 reading skill proficiency (e.g., Harrington & Sawyer, 1992), albeit not with online processing for L2 garden-path sentences in preliminary analyses (Juffs, 2004). Research has also implicated a role for phonological short-term memory differences in L1 word learning and lexical knowledge (Baddeley, Gathercole, & Pagano, 1998), as well as in second language acquisition (N. Ellis, 1996).

Regarding broad language-relevant variables, lexical knowledge (vocabulary) is a significant contributor to reading comprehension abilities in adolescents and adults (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Braze, Tabor, Shankweiler, & Mencl, 2007), making it a relevant variable to account for in our study of college-aged participants. Print exposure, in turn, has been reported to be a significant predictor of lexical knowledge, even after controlling for working memory, age, and education differences (Stanovich, West, & Harrison, 1995; West, Stanovich, & Mitchell, 1993). More generally, print exposure and lexical knowledge can be considered substantial correlates for individuals' amount of reading experience, which may be logically expected to contribute to differences in reading skill. The inclusion of these two measures is therefore of potential importance in assessing the specific contribution of differences in statistical learning skills to language comprehension variance in our sample.

Lastly, we incorporated two nonverbal variables in our design: fluid intelligence and cognitive motivation. Although it has been suggested that

AGL is largely independent of intelligence (e.g., Reber, 1993), measures of fluid intelligence, using a nonverbal test of IQ, have been found to significantly predict individual differences on miniature second-language learning tasks (e.g., Brooks et al., 2006). We therefore included a nonverbal, fluid intelligence measure to test for an association with statistical learning performance in our tasks and to factor this variable out, as necessary, if it correlated with our statistical learning and language measures. Similarly, as motivational differences in our participants' eagerness to be engaged in demanding cognitive tasks (such as the ones employed throughout this experiment) may be a common underlying factor cutting across many of these measures, we measured cognitive motivation to control for this possibility.

Method

Participants and Materials

Thirty monolingual, native English speakers from the Cornell undergraduate population (23 women and 7 men; $M = 19.9$ years, $SD = 1.4$, range = 18-23) participated for course credit or money. To study the relationship between individual differences in statistical learning and language, we administered a test battery assessing two types of statistical learning, language comprehension, lexical knowledge, reading experience, verbal working memory, short-term memory (STM) span, fluid intelligence (IQ), and cognitive motivation. (A summary of the tasks and measures is given in Table 2.1.)

Table 2.1

Descriptive statistics for the individual-differences tasks and measures.

Task	Dependent Measure	Mean (SD)
Statistical learning		
Adjacent	Percent correct (of 40 2AFC items)	62.1 (14.3)
Nonadjacent	Percent correct (of 40 2AFC items)	69.2 (24.7)
Language comprehension		
Animate/inanim. clauses (A/IN)	Percent correct (28 Y/N questions)	90.1 (7.2)
Phonological typicality (PT)	Percent correct (20 Y/N questions)	94.4 (6.7)
Subj./obj. relatives (S/OR)	Percent correct (40 Y/N questions)	85.6 (9.8)
Other language/cognition factors		
Lexical knowledge (SILS)	Number correct (of 40) + (0.25 x number omitted)	34.4 (2.9)
Reading experience (ART)	Proportion correct targets (40) minus checked foils (40)	0.44 (0.16)
Verbal working memory (vWM)	Maximum word span with 2 of 3 trials correct (15 total trials)	4.2 (1.3)
Short-term memory span (FDS)	Number correct trials (of 16) prior to 2 consecutive failures	11.0 (2.3)
Fluid intelligence (CFIT)	Composite raw score (4 subsections, 50 total items)	29.7 (3.6)
Cognitive motivation (NFC)	Sum of scaled responses (ratings for 34 statements)	40.6 (31.6)

Note. A/IN = Animate/Inanimate Noun clauses, PT = ambiguities involving Phonological Typicality, S/OR = Subject-Object Relative clauses, SILS = Shipley Institute of Living Scale, ART = Author Recognition Test, vWM = verbal working memory, FDS = Forward Digit Span, CFIT = Culture Fair Intelligence Test, NFC = Need for Cognition.

Table 2.1 (Continued)

Task	95% Confidence Interval for Mean	Observed Range	Possible Range
Statistical learning			
Adjacent	[56.7, 67.4]	40 - 97.5	0 - 100
Nonadjacent	[60.0, 78.4]	32.5 - 100	0 - 100
Language comprehension			
Animate/inanim. clauses (A/IN)	[87.4, 92.8]	75 - 100	0 - 100
Phonological typicality (PT)	[91.9, 96.9]	72 - 100	0 - 100
Subj./obj. relatives (S/OR)	[81.9, 89.3]	58 - 98	0 - 100
Other language/cognition factors			
Lexical knowledge (SILS)	[33.3, 35.5]	29 - 39	0 - 40
Reading experience (ART)	[0.38, 0.50]	.125 - .725	0 - 1
Verbal working memory (vWM)	[3.7, 4.7]	1.5 - 6	1 - 6
Short-term memory span (FDS)	[10.1, 11.9]	8 - 16	0 - 16
Fluid intelligence (CFIT)	[28.3, 31.0]	19 - 36	0 - 50
Cognitive motivation (NFC)	[28.8, 52.4]	-13 - +108	-136 - +136

Statistical Learning

Two statistical learning tasks, each implementing one of two types of artificial grammars, involving either adjacent or nonadjacent dependencies were conducted. We employed these two types of statistical learning given the possible distinction between these forms suggested by findings and approaches in the literature (see *Introduction*). These types of statistical dependencies also have clear parallels within natural language structure, as sensitivity to adjacent dependencies is important for the discovery of the

relationship between words within phrases and between the phrases themselves (e.g., Saffran, 2001), whereas sensitivity to nonadjacent relationships between constituents is important for embeddings and long-distance dependencies (e.g., Gómez, 2002). Moreover, it has recently been suggested that different brain systems may be involved in the learning of adjacent and nonadjacent dependencies, with only the latter relevant for language (Friederici et al., 2006).

The auditory stimuli and design structure for the statistical learning tasks were typical of those successfully used in the literature to assess statistical learning (e.g., Gómez, 2002). In particular, stimuli strings were constructed by combining individual nonword tokens recorded from a trained female, native English speaker. Assignment of particular tokens (e.g., *pel*) to particular stimulus variables (e.g., the *c* in *cXf* for the nonadjacent statistical learning task, see further below) was randomized for each participant to avoid learning biases due to specific sound properties of words. Nonwords were presented with a 250 msec inter-stimulus interval (ISI) within strings and a 1000 msec ISI between strings.

For both tasks, training lasted about 25 minutes and was followed by a 40-item test phase. Prior to training, participants were informed that they should pay attention to the auditory sequences since they would later be tested on them, but no allusion was made to the existence of any regularities or patterns. After training, participants were informed that the sequences they just heard had been generated according to rules specifying the ordering of the nonwords. They then completed a two alternative forced choice (2AFC)

test in which they were requested to discriminate grammatical strings from ungrammatical ones, with the encouragement to use ‘gut instinct’ and impressions of familiarity/unfamiliarity to guide their judgments. Test-item pairs were presented within two blocks that counterbalanced the presentation order of grammatical and ungrammatical string versions. Half of the test pairs contained novel components that required the participant to be able to generalize the appropriate regularities to new material. These consisted of novel strings for the adjacent statistical learning task and familiar dependency pairs with novel middle elements for the nonadjacent statistical learning task. The other half of test pairs required the participant to recognize previously heard material. These involved the exact strings presented during training. Ungrammatical strings for all test-pair items differed from grammatical ones by only one element.

$$\begin{array}{ll}
 S \rightarrow AP\ BP & S \rightarrow \begin{Bmatrix} a\ X\ d \\ b\ X\ e \\ c\ X\ f \end{Bmatrix} \\
 AP \rightarrow \begin{Bmatrix} D\ A \\ E\ C\ A \end{Bmatrix} & X \rightarrow \{X_1, X_2, \dots, X_{24}\} \\
 BP \rightarrow B\ (AP) &
 \end{array}$$

Figure 2.1. The two artificial grammars used to assess statistical learning of adjacent (left) and nonadjacent (right) dependencies.

For the adjacent statistical learning task, the grammar was adapted with minor modification from Friederici, Steinhauer, and Pfeifer (2002) and contained adjacent dependencies occurring both within and between phrases (see Figure 2.1, left). Regarding phrase internal dependencies, a *D* constituent

always perfectly predicted and occurred prior to an *A* constituent, whereas an *E* constituent always directly preceded a *C* constituent that, in turn, occurred before an *A* constituent (i.e., *E C A*). Between-phrase dependencies resulted from every *B* phrase (*BP*) being consistently preceded by an *A* phrase (*AP*) and optionally followed by another *A* phrase. The language was instantiated through 10 distinct nonword tokens (*biv*, *dupp*, *hep*, *jux*, *lum*, *meep*, *rauk*, *sig*, *tam*, *zoet*) distributed over these lexical categories such that there were 3 *A* members, 3 *B* members, 2 *C* members, 1 *D* member, and 1 *E* member. From a set of 270 unique strings belonging to the grammar, a subset of 60 was selected as training material common to all participants and was presented in three blocks. Ungrammatical strings were produced by replacing a nonword in the string with another from a different category. For instance, if the grammatical string involved the following sequence of category constituents, *D A B D A*, a violation could entail a replacement of the second *D* with an *A*, yielding the ungrammatical string, **D A B A A* (e.g., “*jux hep lum jux biv*” vs. “*jux hep lum hep biv*”). The position of the ungrammaticality was distributed equally across categories with the exception that no violations occurred at the first or last nonword of a string (because such violations are easy to detect; Knowlton & Squire, 1996). Although strings were constructed by selecting nonwords from categories, it is important to point out that participants were exposed to all possible adjacent dependencies during familiarization. Therefore, significant discrimination by participants would reflect knowledge of adjacent structure.

Regarding nonadjacent dependencies, the ability to track relationships among remote dependencies is a fundamental linguistic ability. Previous work

has shown that the statistical learning of nonadjacent dependencies is facilitated in infants and adults when there is high variability in the material that comes between the dependent elements (Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2003; Onnis, Monaghan, Christiansen, & Chater, 2004). We capitalized on this by only exposing learners to a nonadjacent dependency language incorporating high variability. Thus, for the nonadjacent statistical learning task, the grammar conformed to that of Gómez's (2002) high-variability language and consisted of 3 sets of dependency pairs (that is, *a-d*, *b-e*, *c-f*), each separated by a middle *X* element (see Figure 2.1, right). The string-initial (*a*, *b*, *c*) and final (*d*, *e*, *f*) elements that comprise the nonadjacent pairings were instantiated with monosyllabic nonwords (*dak*, *pel*, *vot*; *jic*, *rud* and *tood*). The intervening *X*s were drawn from 24 distinct disyllabic nonwords (*balip*, *benez*, *chila*, *coomo*, *deecha*, *feenam*, *fengle*, *gensim*, *gople*, *hifam*, *kicey*, *laeljeen*, *loga*, *malsig*, *nilbo*, *plizet*, *puser*, *roosa*, *skiger*, *suleb*, *taspu*, *vamey*, *wadim*, and *wolash*). All 72 unique sentences generated from this grammar were presented through six blocks of training. Ungrammatical strings were produced by disrupting the nonadjacency relationship with an incorrect element, thus producing strings of the form **aXe*, **bXf*, and **cXd*.

Language Comprehension

Significant differences can be found in healthy adults' ability to process sentences (see e.g., Farmer et al., in press, for a review). We used a self-paced reading task to investigate the degree to which individual differences in language comprehension are associated with individual differences in

statistical learning performance. Sentences were presented individually on a monitor using a standard word-by-word, moving window paradigm (cf. Just, Carpenter, & Woolley, 1982) and followed by “yes/no” questions probing for comprehension accuracy. While reading times were recorded, the measures of interest for our analyses were the comprehension scores that served as offline correlates of language ability.³ The sentence material consisted of sentences drawn from three different prior studies of various aspects of language processing (see Table 2.2) and chosen for this study because they entail the tracking of adjacent and/or nonadjacent dependencies in natural language. Thus, the sentence set involving clauses with animate/inanimate noun constructions (abbreviated herein as “A/IN”; Trueswell, Tanenhaus, & Garnsey, 1994) contained both adjacent dependencies, i.e., between the animate or inanimate main clause object-noun and its modifying relative clause (e.g., *defendant/evidence [...]RC*), as well as nonadjacent dependencies holding across the relative clause, between the object-noun and the main verb (e.g., *defendant/evidence [...]RC turned*). The sentence set involving noun/verb homonyms with phonologically typical or atypical noun/verb resolutions (abbreviated herein as “PT”; Farmer, Christiansen, & Monaghan, 2006) required tracking adjacent relations between the sentence’s ambiguous homonym and the material that immediately follows it and locally resolves

³ Because typical AGL test-measures of statistical learning, as used here, involve a substantial meta-cognitive component for participants’ offline judgments, we considered these offline language comprehension scores to be a suitable measure for comparisons across the two tasks. Additionally, comprehension accuracy for our participants was below ceiling, with significant variation in performance to serve as an appropriate individual-differences measure.

Table 2.2

The three language comprehension sets, with corresponding examples for each version of a given target sentence.

<p>Subject-Object Relative Clauses (S/OR)</p> <p><i>Subject relative:</i> The reporter that attacked the senator admitted the error.</p> <p><i>Object relative:</i> The reporter that the senator attacked admitted the error.</p> <p>Animate-Inanimate Noun Clauses (A/IN)</p> <p><i>Animate reduced/[unreduced]:</i> The defendant [who was] examined by the lawyer turned out to be unreliable.</p> <p><i>Inanimate reduced/[unreduced]:</i> The evidence [that was] examined by the lawyer turned out to be unreliable.</p> <p>Ambiguities involving Phonological Typicality (PT)</p> <p><i>Noun-like homonym with noun/verb resolution:</i> Chris and Ben are glad that the bird perches [seem easy to install]/[comfortably in the cage].</p> <p><i>Verb-like homonym with noun/verb resolution:</i> The teacher told the principal that the student needs [were not being met]/[to be more focused].</p>
--

the ambiguity (e.g., *bird perches_{homonym} seem_{verb} vs. bird perches_{homonym} comfortably_{adverb}*). The sentence set with subject-object relative clauses (abbreviated herein as “S/OR”; Wells, Christiansen, Race, Acheson, & MacDonald, 2009) required tracking both complex nonadjacent relationships (e.g., between the head-noun and the matrix verb across the embedded clause; *reporter [...]_{RC} admitted*) and relatively simpler, more adjacent relationships (e.g., between the embedded noun and embedded verb; *senator attacked*).

Four sentence lists were prepared, each incorporating 12 initial practice items, 40 sentences with subject-object relative clauses (S/OR), 28 sentences involving clauses with animate/inanimate noun constructions (A/IN), and 20

sentences involving noun/verb homonyms with phonologically typical or atypical noun/verb resolutions (PT). Sentence versions for each target sentence were counterbalanced across the four lists and presented in random order. A comprehension question was presented after each sentence. For example, after reading the last word of the sentence *“The defendant examined by the lawyer turned out to be unreliable,”* the participant would press a “GO” key, which would present a new screen display with the question *“Did the lawyer question the defendant?”*. After recording their response to the question by pressing either the “yes” or “no” key, participants would receive a new sentence and subsequent comprehension probe. Each participant was randomly assigned to a sentence list, and their comprehension accuracy was computed for each set of materials: S/OR, A/IN, and PT.

Lexical Knowledge

As a broad index of language skill spanning across our participants, the Shipley Institute of Living Scale (SILS) Vocabulary Subtest (Zachary, 1994), a standardized measure based on nationally representative norms, was used to assess lexical knowledge, or vocabulary. It is a paper-and-pencil measure consisting of 40 multiple-choice items in which the participant is instructed to select from among four choices the best synonym for a target word. Participants had to complete the task within 10 minutes.

Reading Experience

Measures of print exposure, as intended indicators of reading experience,

have been found to be a significant predictor of individual differences relevant to sentence comprehension, such as vocabulary and orthographic processing (Stanovich & West, 1989; Stanovich et al., 1995). We thus used the Author Recognition Test (ART; Stanovich & West, 1989) as a traditional proxy measure of relative reading experience to assess the extent to which this may account for variance in our participants' language comprehension scores. The questionnaire required participants to check off the names of popular authors on a list. The names belonging to popular writers were chosen from a variety of print media and genres, avoiding standard school curriculum authors. The list was updated from its original form and included 40 actual authors and 40 foils. Two effort probes (the names *Edgar Allen Poe* and *Stephen King*) were also included within the list to check for attentiveness in completing the questionnaire, as these are author names that should be recognized by contemporary monolingual students attending an American college or university.

Verbal Working Memory

Differences in verbal working memory have been associated with individual variations in sentence processing abilities (see MacDonald & Christiansen, 2002, for a review). To determine the degree to which performance on our statistical learning tasks can explain variations in sentence processing skill over and above individual differences in verbal working memory, we used the Waters and Caplan (1996) reading span task as an assessment of our

participants' verbal working memory (vWM).⁴ Participants formed yes/no semantic plausibility judgments for sets of sentences, presented one by one. At the end of a set, participants had to recall all sentence-final words in that set. The number of sentences in each set increased incrementally from 2 to 6, with three trials at each level. Reading span was defined as the maximum level at which a participant correctly recalled all sentence-final words in 2 out of 3 trials, with no more than one failed trial at each of the preceding levels and with half-a-point added if one trial had been correct at the next highest level.

Short-Term Memory Span

Whereas the above-mentioned span task is designed to measure verbal working memory relevant for language processing, we also included an auditory Forward Digit Span (FDS) task, derived from the standardized WAIS-R subtest (Wechsler, 1981), to measure rote memory span. Among psychometric measures of individual differences in verbal short-term memory, the auditory digit span is the most widely used in the literature (Baddeley et al., 1998). A recording played a sequence of digits spoken in monotone at 1-sec intervals. A standard tone after each sequence cued the participant to repeat out loud the digits they had heard in their proper order. Sequences progressed in length from 2 to 9 digits, with two distinct sequences given for each level. Similar to WAIS-R scoring procedures, the dependent measure was the number of correctly recalled trials prior to failure on two

⁴ The Waters and Caplan version was used since it was reported to have greater test-retest reliability than the original Daneman-Carpenter measure (Waters & Caplan, 1996).

consecutive trials.

Fluid Intelligence

General intelligence is another factor that has been suggested to affect individual differences in language and cognition (e.g., Dionne, Dale, Biovin, & Plomin, 2003). Moreover, Brooks et al. (2006) recently found that scores from the Culture Fair Intelligence Test predicted successful learning on an artificial language learning task in many ways similar to our statistical learning tasks. We therefore included this IQ test as a measure of individual differences in intelligence. We used Scale 3, Form A of the Culture Fair Intelligence Test (CFIT; Cattell, 1971), which is a nonverbal test of fluid intelligence or Spearman's (1927) *g*. The test contained four individually timed subsections (Series, Classification, Matrices, Typology), each with multiple-choice problems progressing in difficulty and incorporating a particular aspect of visuospatial reasoning. Raw scores on each subtest were summed together to form a composite score, which may also be converted into a standardized IQ.

Cognitive Motivation

As there may be differences across our participants in their cognitive motivation, we gauged such differences using the Need for Cognition (NFC) Questionnaire (Cacioppo & Petty, 1982) and intended to factor these out in our analyses. The NFC questionnaire provided a scaled quantification of participants' predisposition to engage in and enjoy effortful cognitive activities. Participants indicated the extent of their agreement/disagreement

to 34 particular statements (e.g., “*I prefer life to be filled with puzzles that I must solve*”). We planned to examine how this measure correlates with language and statistical learning and to incorporate it as a covariate if necessary.

Procedure

Participants were individually administered the tasks during two sessions conducted on separate days (within a span of 2 - 9 days apart; mean interval = 5.2 days, $SD = 2.0$). For each participant, one of the two statistical learning tasks was randomly assigned for the beginning of the first session, and the other was given at the start of the second session. The remaining tasks were divided into two sets with fixed order. Set A consisted of the self-paced reading task, followed by the SILS vocabulary assessment, the NFC, and then the FDS; Set B consisted of the CFIT, the vWM span task, and then the ART. Each participant was randomly assigned one of these sets (A or B) for the first session, with the other set administered during the second session.

Results

The means, standard deviations, and range for all measures are provided in Table 2.1. Average performance on the two statistical learning tasks—62.1% ($SD=14.3\%$) and 69.2% ($SD=24.7\%$) for adjacent and nonadjacent materials,⁵

⁵ Since 5 participants received an erroneous 2AFC test-pair item on the nonadjacent statistical learning task (prior to the test-pair being corrected during the course of the experiment), scores are reported as the proportion correct (with the erroneous test-pair item removed for the affected individuals). The erroneous test-pair contained two ungrammatical test strings (* $a X_7 e$ vs. * $b X_7 d$) before being corrected to * $a X_7 e$ versus $b X_7 e$. None of the other test-pair items consisted of any of these specific strings (i.e., * $a X_7 e$, * $b X_7 d$, and $b X_7 e$).

respectively—was significantly above chance-level classification and indicative of learning at the group-level; $t(29) = 4.63, p < .0001$ for the adjacent statistical learning task; $t(29) = 4.26, p = .0002$ for the nonadjacent statistical learning task. Each of the statistical learning tasks contained a balanced number of generalization and recognition test items (incorporating “novel” and “familiar” components respectively, as detailed under *Methods*). The average gain in accuracy for generalization items compared to recognition items was 1.2% ($SE=2.3$) for the adjacent statistical learning task [matched pairs t -test: $t(29) = .51, p = .61$] and was -0.8% ($SE=2.0$) for the nonadjacent statistical learning task [matched pairs t -test: $t(29) = .39, p = .70$]. As participants did not significantly differ in their performances on generalization and recognition tests, we collapse across these tests in subsequent analyses. Due to the experiment design, some participants received the adjacent statistical learning task during their first session ($n = 18$), whereas others received the nonadjacent statistical learning task first ($n = 12$). However, there was no main effect of statistical learning task order on participants’ statistical learning scores, $F(1,28) < 1, p = .64$.

The first objective in our analyses was to determine the relation between adjacent and nonadjacent dependency learning. Based on whether these correlated significantly, we intended to conduct either partial correlation analyses (in the affirmative case) or standard bivariate analyses (if no correlation was obtained). Using as our central language measures the three language scores derived from the self-paced reading task (i.e., comprehension

subscores, differentiated by sentence-type)⁶, we planned to explore significant correlations found between the three language measures and each of the two statistical learning measures as well as the other individual difference factors, when the effects of all measures other than a given predictor were held constant. We found no correlation between the two statistical learning tasks ($r = .14, p = .45$), as shown in Figure 2.2.

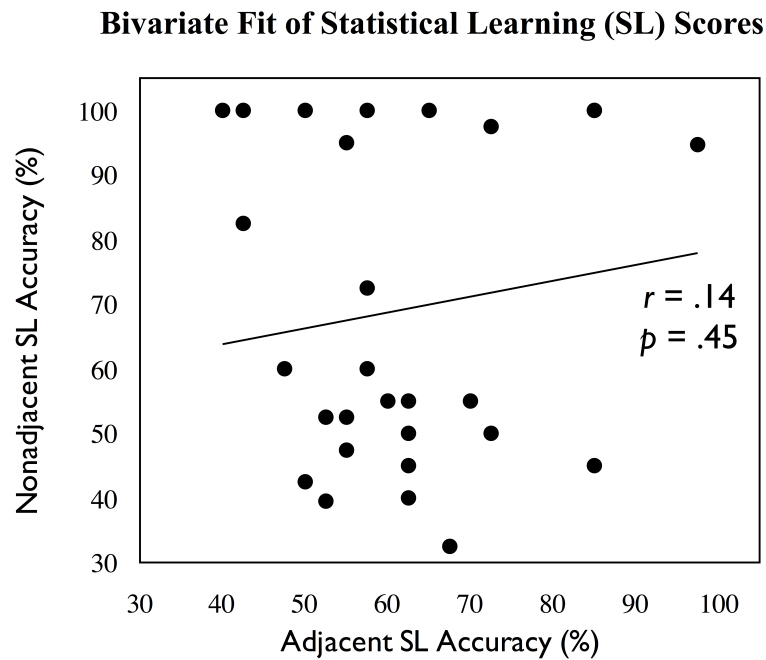


Figure 2.2. Participants' accuracy scores for the adjacent statistical learning (SL) task (x-axis), plotted against their accuracy scores for the nonadjacent SL task (y-axis).

⁶ There were a few coding errors in the programs for presenting some of the sentence lists, resulting in: one fewer presented item in the sentence set corresponding to the PT manipulation for Lists 2 and 3; 4 fewer sentences (1 S/OR item, 1 A/IN item, and 2 PT items) in List 1. List 4 was error-free. However, there was no significant effect of List on comprehension accuracy for the sentence-type sets, $F(3, 26) < 1, p = .42$.

Table 2.3

Intercorrelations between task measures.

	Statistical Learning		Lang. Comprehension			Other Lang./Cognition Factors				
	Adjacent	Nonadj.	A/IN	PT	S/OR	SILS	ART	vWM	FDS	CFIT
NA-SL	.14									
A/IN	-.02	.41*								
PT	.49**	.12	.18							
S/OR	.39*	.42*	.11	.46*						
SILS	.05	.26	.28	.33†	-.07					
ART	-.17	.16	.37*	.14	-.05	.33†				
vWM	.46*	.53**	.37*	.40*	.39*	.35†	.22			
FDS	.40*	.13	.02	.32†	.33†	.11	-.20	.36†		
CFIT	.23	.19	.20	.02	.01	.21	.07	.28	.16	
NFC	.22	.15	.33†	.32†	.03	.34†	.20	.27	.03	-.08

† $p < .09$. * $p < .05$. ** $p < .01$ (two-tailed, $n = 30$).

Note. NA-SL = Nonadj. = Nonadjacent statistical learning, A/IN = Animate/Inanimate Noun clauses, PT = ambiguities involving Phonological Typicality, S/OR = Subject-Object Relative clauses, SILS = Shipley Institute of Living Scale, ART = Author Recognition Test, vWM = verbal working memory, FDS = Forward Digit Span, CFIT = Culture Fair Intelligence Test, NFC = Need for Cognition.

We then computed the correlations between all task measures, as shown in Table 2.3. Regarding statistical learning, adjacent dependency learning (“Adj-statistical learning”) was positively associated with comprehension for the sentence set involving phonological-typicality ambiguities (“PT comprehension”), comprehension for the sentence set involving subject-object relative clauses (“S/OR comprehension”), vWM, and FDS; nonadjacent dependency learning (“Nonadj-statistical learning”) was positively associated with comprehension for the sentence set involving animate-inanimate noun

clauses (“A/IN comprehension”), S/OR comprehension, and vWM. As can be seen in Table 2.3, all statistically significant correlations were of medium size, ranging between .39 and .53.

For the language-processing measures, A/IN comprehension—in addition to the positive correlation with nonadjacent statistical learning noted above—correlated with ART and vWM. PT comprehension, as well as correlating with adjacent statistical learning (above), was further positively associated with S/OR comprehension and vWM. S/OR comprehension—besides correlating with adjacent statistical learning, nonadjacent statistical learning, and PT comprehension—correlated with vWM. Note then that there was considerable overlap in the language correlations obtained between (and among) nonadjacent statistical learning, adjacent statistical learning, and vWM. Additionally, the specific pattern of intercorrelations between statistical learning and vWM/STM indicate that adjacent statistical learning is relatively strongly associated with both vWM and STM performance ($r = .46$ and $.40$, respectively). The vWM measure is also correlated substantially with nonadjacency learning ($r = .53$), whereas STM performance only has a very weak correlation with such statistical learning ($r = .13$, $p > .09$). Thus, the kind of learning and memory skills involved in the vWM task may be more closely related to the learning of nonadjacencies than adjacencies. In contrast, the STM measure may be more closely associated with mechanisms subserving the learning of adjacent dependencies.

Table 2.4

Intercorrelations between language comprehension measures (left column) and statistical learning and memory-span measures (top row).

	Adj.-SL	NA-SL	vWM	FDS
A/IN	-.02	.41*	.37*	.02
PT	.49**	.12	.40*	.32†
S/OR	.39*	.42*	.39*	.33†

† $p < .09$. * $p < .05$. ** $p < .01$ (two-tailed, $n = 30$).

Note. Adj.-SL = Adjacent statistical learning, NA-SL = Nonadjacent statistical learning, vWM = verbal working memory, FDS = Forward Digit Span, A/IN = Animate/Inanimate Noun clauses, PT = ambiguities involving Phonological Typicality, S/OR = Subject-Object Relative clauses.

Table 2.5

Regression coefficients of predictor variables for each of the dependent language measures: A/IN (Animate/Inanimate Noun clauses), PT (ambiguities involving Phonological Typicality), and S/OR (Subject-Object Relative clauses).

	Statistical Learning		Other Predictors					
	Adjacent	Nonadjacent	SILS	ART	vWM	FDS	CFIT	NFC
A/IN	-.20	.25	-.03	.20	.18	.01	.16	.27
PT	.42*	-.08	.24	.14	.12	.15	-.18	.07
S/OR	.28	.38*	-.20	.04	.13	.18	-.17	-.09

* $p < .05$ (one-tailed, $n = 30$).

Note. SILS = Shipley Institute of Living Scale, ART = Author Recognition Test, vWM = verbal working memory, FDS = Forward Digit Span, CFIT = Culture Fair Intelligence Test, NFC = Need for Cognition.

To ease direct comparisons between the statistical learning and memory-related measures, their intercorrelations from Table 2.3 are transcribed more compactly in Table 2.4. As evident in the table, vWM is well correlated with language performance in general, whereas each type of statistical learning appears to be associated more specifically with a subset of the sentence structures (as examined further in the next set of analyses).

To determine how well each measure predicted language comprehension, when controlling for all other predictors, we obtained the regression coefficient values of our individual-differences variables for each dependent language measure (Table 2.5). For A/IN, none of the predictors reached significance. For PT, however, adjacent statistical learning (but none of the other variables) showed a strong positive relationship to language comprehension when all other factors were held constant ($\beta = .42, p < .05$, one-tailed t-test). For S/OR, only nonadjacent statistical learning was strongly related to language comprehension ($\beta = .38, p < .05$, one-tailed t-test).⁷

Notably, the regression coefficients for verbal working memory were weaker (i.e., 0.18 or less) and much farther from reaching significance (all p 's $> .24$). In each case then, when controlling for the effect of all other predictors, the only predictor that makes a significant and substantial contribution to the PT and

⁷ As we had specifically predicted the direction of the correlation between the statistical learning and language comprehension measures to be positive, one-tailed tests were used. However, two-tailed t-tests would still yield marginally significant correlations between adjacent statistical learning and PT comprehension ($p = .0596$) and between nonadjacent statistical learning and S/OR comprehension ($p = .0751$).

S/OR language-processing measures was either of the two statistical learning measures.

Finally, we note that any comparisons between statistical learning and vWM/STM measures are clearly limited by the tasks used to assess them. If the tasks used to assess a construct is poorer than those used to assess the comparative construct(s), then the beta weights for the former predictor in regression analyses could be misleadingly attenuated in relation to the others. In principle, this could be true for vWM, though we used an established reading-span task with high internal consistency and reliability (Waters & Caplan, 1996). In principle, this caveat is also applicable to our statistical learning measures; specifically in that regard, our nonadjacency task may have been prone to a ceiling effect (see Figure 2.2), which may have limited upper-range variability in our sample and potentially reduced what would otherwise have been an even larger beta value.

Discussion

Only a few prior statistical learning studies have reported quantitative differences in performance across participants. First, with regards to adjacent statistical learning, adults performed within the range of 41-83% accuracy in the test phase of a linguistic segmentation task studied by Saffran, Newport, Aslin, Tunick, and Barrueco (1997; as reported in Evans et al., 2009). A similar range is evident upon inspection of the figures in Saffran et al. (1999), wherein the average lower-bound for adult performance on nonlinguistic segmentation tasks is approximately 49% and the average upper-bound is approximately

89%. The lowest and highest performances, respectively, across the multiple studies in Saffran et al. (1999) appear to be 33% and 97%. Although our adjacent statistical learning task involves learning an artificial grammar rather than an artificial lexicon, these observations are nonetheless consistent with our reported performance range from 40 to 97.5%.

Secondarily, with regards to our study's nonadjacent statistical learning task which replicates the design of Gómez's (2002) high-variability condition, Gómez had noted that two-thirds of learners in this condition showed perfect discrimination on a grammaticality-endorsement test measure. Analogously, we also observed perfect to near-ceiling ($\geq 95\%$) performances by eleven learners in this study. Although this is proportionally less than that reported by Gómez, mean performance in the original high-variability condition was also substantially higher, at 90% accuracy, than in subsequent replications where average performance is comparable to the more modest level reported herein (see Van den Bos, Christiansen, & Misyak, 2012). Thus, although there are few documented details regarding statistical learning variation in normal adults, the variance captured by our tasks generally accords with what is known within the standard literature. This study is among the first to methodically record such information, and we encourage future researchers to include such information in their reported results.

While it has been traditionally assumed that statistical learning processes (as commonly studied here, using AGL tasks) are largely invariant across individuals (e.g., Reber, 1993), our findings instead documented systematic variability in statistical learning performance within the normal

adult population. This coincides not only with a recently emerging recognition that individual differences may exist, as even conceded in Reber and Allen (2000), but also with the development of newer paradigms intended to specifically tap into such differences (e.g., Karpicke & Pisoni, 2004; Misyak, Christiansen, & Tomblin, 2010b). As an initial investigation into these differences using a comprehensive within-subject design, our results indicated that statistical learning scores are substantially and reliably interrelated with vWM and language comprehension. Moreover, when controlling for the effects of all other predictors in the regression analyses, statistical learning ability, rather than vWM, was the only predictor of comprehension accuracy for two of the main types of sentence materials. Following MacDonald and Christiansen (2002; see also Wells et al., 2009), these results are consistent with the likely role of vWM as another index of processing skill for language comprehension and statistical learning, rather than a functionally separate capacity or mechanism. Indeed, differences in statistical learning have been recently shown to capture key on-line language processing patterns previously attributed to vWM differences (Misyak et al., 2010a).

Furthermore, the specific pattern of correlations between statistical learning measures and language comprehension subscores suggests that individual differences in detecting adjacent and nonadjacent dependencies may map onto variations in corresponding skills relevant to processing similar kinds of dependencies as they occur in natural language. Thus, comprehending subject-object relative constructions in the S/OR material entails tracking long-distance relationships spanning across lexical

constituents (e.g., relating the object of an embedded clause to the subject and main verb of the sentence). Analogously, the processing of items in the PT set relies upon sensitivity to adjacent information, in which the ambiguous homonym is disambiguated and locally resolved by the next word.

Because of the correlation first reported by Brooks et al. (2006) between CFIT scores and their language-learning task, we had computed the correlations between CFIT and our statistical learning tasks, but did not detect any significant associations. Scores, though, for nearly all our participants were above their reported median and likely comprised a narrower range. Moreover, the lack of any associations may be consistent with subsequent findings by Gebauer and Mackintosh (2007) in which fluid intelligence correlated with artificial grammar learning when participants were given detailed instructions for how to intentionally look for patterns in the training material, but not when participants were administered AGL tasks under typical instructions (as here) which promoted more incidental learning. However, our sample size combined with the probably narrower range of observed CFIT scores may have conversely limited our power to detect any potential associations.

Our experimental design included a battery of other measures that have previously received attention in studies of L1 and L2 language learning, such as lexical knowledge (vocabulary), reading experience, cognitive motivation, and short-term memory span. Among these, lexical knowledge marginally correlated with print exposure (as also replicated by significant findings in Braze et al., 2007), supporting arguments for amount of reading as

the best contributor to vocabulary breadth (De Temple & Snow, 2003; Krashen, 1989; Nagy & Anderson, 1984; Stanovich, 1986). The short-term memory measure (assessed via the FDS task) was found to correlate positively with adjacent statistical learning, but not nonadjacent statistical learning. Karpicke and Pisoni (2004) also had reported a correlation of equal magnitude between auditory digit span and AGL performance on an implicit sequence-learning task involving auditory or auditory-visual stimuli. Thus, the ability to recall successive elements of numerical series may covary with adjacent statistical learning skill, but this finding does not necessarily entail that a parallel relationship exists between short-term memory of phonological sequences and nonadjacency learning.

With the exception of short-term memory span and adjacency learning, then, the fact that the present study did not detect significant, strong interrelationships between most of these other variables and statistical learning performance or language comprehension does not deny their potential importance within accounts of language learning. It is furthermore possible that some of these canonical measures may be relatively weak proxies for their intended constructs (e.g., for discussion of potential limitations associated with using the ART to assess reading experience, see Acheson, Wells, and MacDonald, 2008). However, it does preliminarily support the thesis that individual differences in statistical learning skills themselves, which have been much overlooked in many explanations to date, may account for a larger proportion of language variance than the more standard measures typically used for individual differences research.

Conclusion

Overall, our findings substantiate this study's motivating rationale that variation within the normal population should provide a suitable framework for testing the empirical relatedness of language and statistical learning. As a confirmation of this approach, we found that individual differences in statistical learning exist and that sensitivity to particular kinds of statistical regularities (i.e., adjacent or nonadjacent) in the artificial grammars are predictive of processing ability for different types of sentence constructions (i.e., involving the tracking of either local or long-distance relationships). Admittedly, our study is limited by the correlational nature of its design, which cannot reveal causality, and by the relatively low number of participants, which reduces statistical power. Nonetheless, the significant results obtained here are encouraging, and should be followed up by a larger-scale study incorporating structural equation modeling to test these hypothesized relationships. Importantly, these findings begin to establish a heretofore-missing empirical link within individuals between statistical learning and language processing (see also Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak et al., 2010a, 2010b).

Our results may also have wider theoretical relevance to questions regarding the nature of underlying mechanism(s) for statistical learning. Although group performances for adjacent and nonadjacent grammar tasks have been documented, the research presented here is the first to assess within-subject differences across these tasks. The lack of any significant correlation detected between them, and possibly the differentiation of their

predictive relations to the language measures, raises an intriguing question as to whether the two types of statistical learning may be subserved by separate mechanisms (see also, Friederici et al., 2006). However, it is also possible that differences in learning strategies or task demands across the two tasks may explain the lack of association between adjacency and nonadjacency learning. Potential bimodality in the distribution of nonadjacency scores may also contribute to the lack of association. If so, this concern might be addressed with future work using newer tasks that more sensitively assess statistical learner differences (e.g., as in Misyak et al., 2010b, in which a fairly continuous and normal distribution of nonadjacency differences was documented).

The overall pattern of findings is consistent with an overlap among underlying mechanisms for both types of statistical learning and those involved for language. Counter to the claims of Friederici et al. (2006), then, we have also found that adjacency learning is substantially implicated in language for the types of sentence structures studied here. More broadly, the notion that statistical learning abilities and language may share common neurocognitive mechanisms also converges with other recent neural evidence (see, e.g., Christiansen, Conway, & Onnis, 2007; Conway & Pisoni, 2008; Friederici et al., 2002; Petersson, Forkstam, & Ingvar, 2004) and with the behavioral findings relating group differences in statistical learning to language-impaired populations (as noted within the *Introduction*). More research that, as here, makes within-subject comparisons across tasks is needed to understand the proper relation between different types of statistical learning and the degree to which they may be relying on the same or different

neural underpinnings. Future work examining individual differences in language and statistical learning should thus aim to study in more detail the relationship between specific types of statistical structure and linguistic processing, while elucidating the nature of the underlying mechanisms upon which statistical learning and language may commonly supervene.

REFERENCES

- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278-289.
- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, 14, 464-473.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.
- Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, 24, 119-131.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382-1407.
- Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96-151). Oxford: Basil Blackwell.
- Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, 40, 226-243.
- Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*,

27, 185-209.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2007). Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 173-178). Austin, TX: Cognitive Science Society.

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, 116, 382-393.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235-253.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114, 356-371.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24-39.

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, 1145, 113-131.

- De Temple, J., & Snow, C. E. (2003). Learning words from books. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Teachers and parents* (pp. 16-36). Mahway, NJ: Lawrence Erlbaum Associates.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1355-1367.
- Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development*, 74, 394-412.
- Don, A. J., Schellenberg, E. G., Reber, A. S., DiGirolamo, K. M., & Wang, P. P. (2003). Implicit learning in children and adults with Williams syndrome. *Developmental Neuropsychology*, 23, 201-225.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91-126.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525-551). Oxford: Blackwell.

- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321-335.
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, 103, 12203-12208.
- Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (in press). *Individual differences in sentence processing*. To appear in: M. Spivey, M. Joannisse, & K. McRae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge, UK: Cambridge University Press.
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458-467.
- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwender, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, 103, 2458-2463.
- Friederici, A.D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, 99, 529-534
- Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence

- dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 34-54.
- Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16, 486-490.
- Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249-268.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Gómez, R. L. (2007). Statistical learning in infant language development. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
- Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178-186.
- Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, 7, 567-580.
- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183-206.
- Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science*, 18, 254-260.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*,

39, 158-170.

- Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25-38.
- Hsu, H., Tomblin, J. B., & Christiansen, M. H. (2009) Statistical learning of nonadjacent dependencies in adolescents with and without specific language impairment. Manuscript in preparation.
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658-680.
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102, 199-225.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Karpicke, J., D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*, 32, 956-964.
- Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, 8, 43-52.
- Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, 58, 703-746.
- Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, 60, 127-153.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on

- implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169-181.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73, 440-464.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35-54.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11, 122-134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-B111.
- McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 239-245.
- Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual: Individual differences and consequences for language production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.
- Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link

- between statistical segmentation and word learning in adults. *Cognition*, 108, 271-280.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1:31. doi: 10.3389/fpsyg.2010.00031
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138-153.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19, 304-330.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. L. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society* (pp. 886-891). Boston, MA: Cognitive Science Society.
- Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1047-1052). Chicago, Illinois: Cognitive Science Society.
- Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, 76, 324-

- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, 130, 401-426.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233-238.
- Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, 28, 383-407.
- Plante, E., Gómez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35, 453-462.
- Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia*, 10, 61-76.
- Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227-247). Amsterdam: John Benjamins Publishing Company.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 17, 888-896.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, 27, 235-268.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101-105.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, 4, 273-284.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.

- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24, 402-433.
- Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, 31, 811-826.
- Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1131-1142.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318.
- Van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). *Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration*. Submitted manuscript.
- Vasilyeva, M., Waterfall, H., & Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Developmental Science*, 11, 84-97.
- Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 387-398.
- Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1289-1295.

- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49, 51-79.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250-271.
- Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.
- West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, 28, 34-50.
- Zachary, R. A. (1994). *Shipley Institute of Living Scale, Revised Manual*. Los Angeles: Weston Psychological Services.

CHAPTER 3

When Better Statistical Learning Leads to Poorer Language Processing

With the expansion of studies on statistical learning over the past decades, focus has intensified towards probing the potential role for probabilistic sequence learning in acquiring and using linguistic structure (e.g., Gómez, 2002; Saffran, 2001). A clearer understanding has in turn begun to crystallize about the ways in which statistical learning mechanisms may underpin language across various levels of organization—phonetic, lexical, semantic, syntactic—and across differing timescales—phylogenetic, ontogenetic, and microsecond unfoldings. Largely missing from this picture, however, has been empirical evidence that directly links language and statistical learning abilities within the typical population.

Thus, a general assumption of current statistical learning research is that statistical learning and language processes are interrelated, with positive correspondence in intraindividual variation across them. But is it always the case that greater statistical learning should be linked with better language functioning? Or, may excelling at one of these implicate poorer performance at the other? Such ability-linked reversals in cognitive performance are not unprecedented. For example, bilingual individuals appear to possess more efficient ‘inhibitory control’ processes than their monolingual peers, which is usually imputed to bilinguals’ greater experience with ‘control’ processes for suppressing irrelevant information in the course of successfully using two languages (see Bialystok, Craik, Klein & Viswanathan, 2004). However, an

opposite pattern was obtained in a negative priming paradigm whereby previous distractors became relevant for facilitating responses to a current trial. The expected facilitation effect was observed for monolinguals but not for bilinguals, whose performance accuracy decreased from a neutral baseline (Trecanni, Argyri, Sorace & Della Sala, 2009). Analogously then, might there be natural language contexts in which superior statistical learning skill also becomes disadvantageous?

Much statistical learning research to date has investigated humans' sensitivity to adjacent statistical cues in distributional input. As early as a few days old, newborns appear sensitive to co-occurrence frequencies occurring within a sequence of geometric shapes or speech sounds (Bulf, Johnson, & Valenza, 2011; Teinonen, Fellman, Näätänen, Alku, and Huotilainen 2009). By two months, infants also evince sensitivity to bigrams, or first-order adjacent pairs, that are identifiable from the co-occurrence frequencies of elements within a constrained temporal sequence (Kirkham, Slemmer & Johnson, 2002). Throughout later development and adulthood, humans can use adjacent conditional probabilities to locate relevant constituent-boundaries in a continuous stream composed of nonwords, tones, visual elements, or nonlinguistic sounds (e.g., Fiser & Aslin, 2002; Gebhart, Newport & Aslin, 2009; Saffran, Newport, Aslin, Tunick & Barrueco, 1997; Saffran, Johnson, Aslin & Newport, 1999). And further, adjacency learning—both in terms of adjacent predictive dependencies (Saffran, 2001) and transitional probabilities between adjacent words (Thompson & Newport, 2007)—has been shown to help with acquiring the underlying structure of an artificial language. But it

remains unclear how statistical learning of adjacent dependencies empirically relates to the processing of more complex natural language structure, which characteristically involves longer-distance dependencies.

One possibility is that a statistical learner may focus too much on computing certain statistics, while ignoring others, with repercussions for linguistic processing. Thus, although the ability to track adjacent statistical dependencies can scaffold the detection of more remote relations (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007), too much sensitivity to such adjacencies may interfere with the learning of long-distance dependencies. Indeed, the processing of locally coherent structures in natural language have been shown to hamper the processing of longer-distance dependencies when the former conflicts with the correct global interpretation of a sentence (Tabor, Galantucci, & Richardson, 2004).

To investigate the possible interrelationship between individual differences in statistical learning of adjacent dependencies and language, we introduce a new AGL-SRT task that incorporates the structured, probabilistic input of artificial grammar learning (AGL; Reber, 1967) within a modified two-choice serial reaction-time (SRT; Nissen & Bullemer, 1987) layout, using auditory-visual sequence-strings as input. Experiment 1 thus documents the group trajectory and range of individual differences for adjacency learning obtained from this task. Test scores reflecting individual differences in adjacency learning are then used in Experiment 2 to probe relationships to natural language processing patterns.

Experiment 1: Individual Differences in Statistical Learning of Adjacencies

While the ability to track adjacent sequential relationships is a widely-acknowledged robust phenomenon (cf. Perruchet & Pacton, 2006), only a few studies have aimed to identify systematic individual differences in such incidental abilities within the normal population of adults (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kaufman et al., 2010; Misyak & Christiansen, 2012; Pretz, Totz, & Kaufman, 2010), young children (Kidd, 2012), or infants (Shafro, Conway, Field & Houston, 2012). Furthermore, the empirical relationship between adjacency learning and on-line syntactic processing remains unknown. We used the biconditional grammar of Jamieson and Mewhort (2005) to examine individual differences in adults' statistical learning of adjacent dependency pairs. This grammar is defined by first-order adjacent transitions only, imposes no positional constraints on element placement, and generates strings of equal length, thus allowing us to isolate the learning of predictive adjacencies.

Method

Participants

Thirty native English speakers from the Cornell undergraduate population (15 females; age: $M=19.4$, $SD=0.8$) were recruited for course credit.

Materials

Participants observed sequences of auditory-visual strings generated by an eight-element grammar in which every element could be followed by one of

only two other elements, with equal probability. Each string consisted of 4 elements, with adjacent probabilities between them as shown in Table 3.1.

Table 3.1

Transition probabilities for elements at positions n and $n + 1$ of a string, with n as an integer from (0, 4).

Element at n	Element at position $n + 1$ of string							
	a	b	c	d	e	f	g	h
a	0	.5	.5	0	0	0	0	0
b	0	0	.5	.5	0	0	0	0
c	0	0	0	.5	.5	0	0	0
d	0	0	0	0	.5	.5	0	0
e	0	0	0	0	0	.5	.5	0
f	0	0	0	0	0	0	.5	.5
g	.5	0	0	0	0	0	0	.5
h	.5	.5	0	0	0	0	0	0

The stimulus nonwords (*jux*, *tam*, *hep*, *sig*, *nib*, *cav*, *biff*, and *lum*) were randomly assigned to the element tokens (a , b , c , d , e , f , g , h) for each participant to avoid potential learning biases due to specific sound properties of words. Auditory versions of the nonwords were recorded from a female native English speaker and length-edited to 490 ms. Written versions of nonwords were presented with standard spelling in Arial font (all caps) and

appeared within the rectangles of a 2 x 4 computer grid (see Figure 3.1). Each of the 4 columns of the computer grid, from left to right, displayed the nonword options corresponding to the 1st through 4th respective elements of a string. Ungrammatical strings were created by introducing an incorrect element at the 2nd or 3rd string position, with the next element being one that legally followed the incorrect one (e.g., as in “a *d e g”).

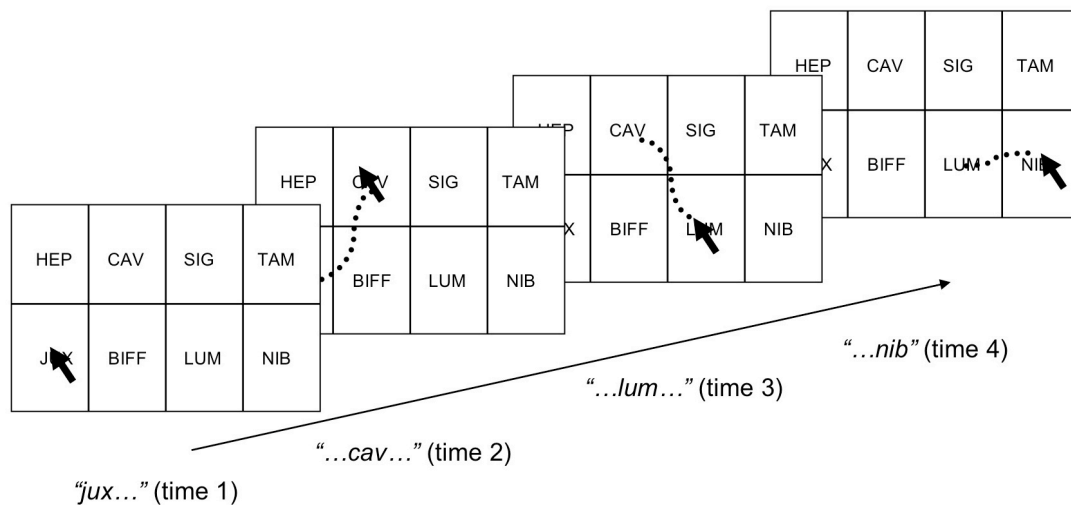


Figure 3.1. The pattern of mouse clicks for a single trial with the auditory target string “jux cav lum nib.”

Procedure

Each trial corresponded to a different configuration of the grid, with each of the eight written nonwords centered in one of the rectangles. Every column contained two nonwords, a target (from a stimulus string) and a foil. The first column contained the selection for the first element of a string, the second column contained the selection for the second element, and so on. For example, a trial with the stimulus string *jux cav lum nib*, as shown in Figure

3.1, might contain the target *jux* and the foil *hep* in the first column; the target *cav* and the foil *biff* in the second column; the target *lum* and the foil *sig* in the third column; and the target *nib* and the foil *tam* in the fourth column. Each nonword appeared equally often as target and as foil within and across the columns. The top and bottom locations of targets and foils were randomized and counterbalanced.

Participants were informed that the purpose of the grid was to display their selections and that a computer program randomly determines a target's location within either the top or bottom rectangle. On every trial, participants heard an auditory stimulus string composed of four nonwords and were instructed to respond to each nonword in the sequence as quickly and as accurately as possible by using the computer mouse to select the rectangles displaying the correct targets. Thus for any given trial, after 250 ms of familiarization to the visually presented nonwords, the first nonword of a string (the target) was played over headphones. Next, the second, third, and fourth words of a given string were each played after a participant had responded in turn to the prior nonword. The grid lines associated with each respective column of the grid became lightly bolded in turn to assist participants in selecting each nonword. For example, on a trial with the stimulus string *jux cav lum nib*, the participant should first click the rectangle containing JUX upon hearing *jux* (Fig. 3.1, left), CAV upon next hearing *cav* (Fig. 3.1, center-left), LUM upon hearing *lum* (Fig. 3.1, center-right), and NIB upon hearing *nib* (Fig. 3.1, right). After a participant had responded to the last nonword, the screen cleared for 750 ms before a new trial began.

An intended consequence of this design is that, for any given trial, the first element of a string cannot be anticipated in advance of hearing the auditory target. However, all subsequent string transitions might be reliably anticipated using statistical knowledge of their bigram structure (i.e., first-order adjacent relations). Thus, as participants become sensitive to the adjacent dependencies, they should be able to anticipate the string transitions, which should be evidenced by faster response times (following standard SRT rationale for assessing learning; e.g., Nissen & Bullemer, 1987). Accordingly, our dependent measure on each trial was the reaction time (RT) for a predictive target, subtracted from the RT for the non-predictive initial-column target (which serves as a baseline and controls for practice effects). The predictive target used in this calculation was equally distributed across all non-initial columns across trials. Analogously, for an ungrammatical string trial, if participants are sensitive to the adjacent dependencies, then their RTs for incorrect, or violated, elements should be slower; thus, the DV for ungrammatical trials was the RT for the illegal target subtracted from the initial-target RT.

There are 64 unique strings ($8 \times 2 \times 2 \times 2$) defined by the grammar; these were all randomly presented once in each grammatical block of trials. Training consisted of six grammatical blocks, followed by an ungrammatical block of 16 trials and then a single grammatical ('recovery') block. Transitions across blocks were seamless and unannounced.

After these eight blocks, participants were informed that the strings had been generated according to rules specifying the ordering of nonwords and

were asked to complete an adjacency pair test. Participants were randomly presented with 32 test items of auditory nonword-pairs. They were requested to judge whether each pair followed the rules of the grammar by pressing 'yes' / 'no' computer keys. Half of the test items were the 16 adjacent transitions licensed by the grammar (e.g., *a b*); the remaining half were illegal pairings formed by reversing each adjacent transition (e.g., *b a*). Thus, successful discrimination reflects knowledge of the conditional dependencies, rather than only sensitivity to co-occurrences.

Results and Discussion

Analyses were performed only on accurate string-trials with a single selection for each target. Prior to analysis, the data from five participants were omitted (2 for withdrawing participation; 2 for improperly performing the task, with less than 40% accurate trials; and 1 for abnormally elevated RTs, averaging in excess of 1470 ms per single response). For remaining participants, accurate trials averaged 88.2% ($SD = 5.9$) of training block trials.

Mean RT difference scores, as described above (i.e., for grammatical trials: initial-target minus predictive-target RT; for ungrammatical trials: initial-target minus illegal-target RT) were computed for each block and submitted to a one-way repeated-measures analysis of variance (ANOVA) with block as the within-subjects factor. Since the assumption of sphericity was violated ($\chi^2(27) = 113.27, p < .001$), degrees of freedom were corrected using Greenhouse-Geisser estimates ($\epsilon = .33$). Results indicated a main effect of block on RT difference scores, $F(2.31, 55.36) = 3.82, p = .02$. As seen in

Figure 3.2, mean RT difference scores appear to increase by the final training block, decrease in the ungrammatical block, and increase once again in the recovery block. As RT difference scores measure the amount of facilitation from the predictive targets, an improvement in scores across blocks (as seen here) reflects an acquired sensitivity to the adjacent dependencies.

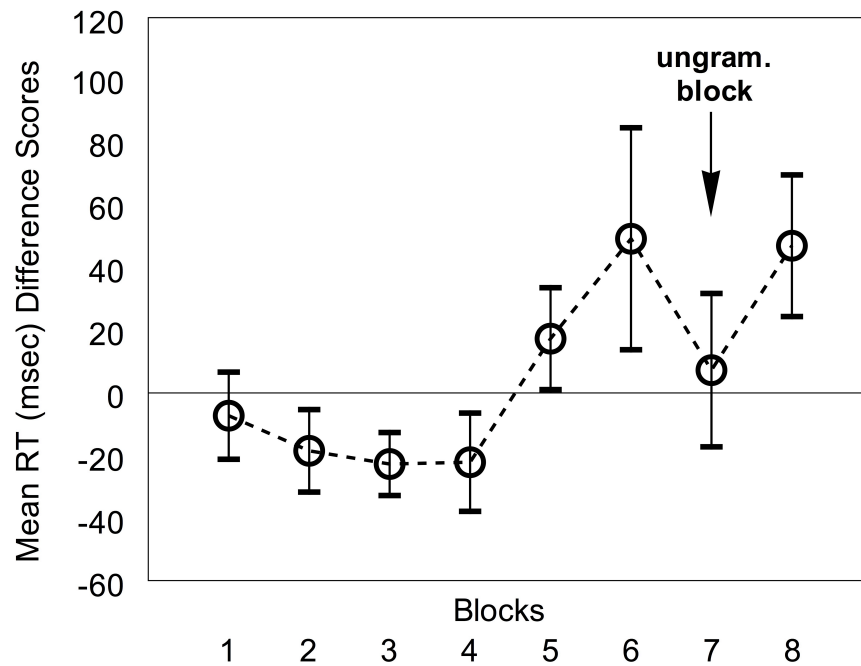


Figure 3.2. Group learning trajectory as a plot of mean reaction time (RT) difference scores per block (for grammatical trials: initial-target minus predictive-target RT; for ungrammatical trials: initial-target minus illegal-target RT). Error bars indicate standard errors.

Planned contrasts between the ungrammatical block and preceding/succeeding grammatical blocks confirmed a performance decline for the ungrammatical trials (Block 6 minus Block 7: $M = -42.0$ ms, $SE = 19.6$, $t(24) = 2.14$, $p = .04$; Block 8 minus Block 7: $M = 39.8$ ms, $SE = 17.8$ ms, $t(24) =$

2.23, $p = .04$). This provides evidence for participants' learning of the sequential dependencies, consistent with standard interpretations in the sequence learning literature for comparing RTs to structured versus unstructured material (e.g., Thomas and Nelson, 2001).

Accuracy on the adjacency pair test also reflected statistical adjacency learning ($t(24) = 4.66$, $p < .0001$), with a mean of 57.6% ($SD = 8$), and scores ranging from 37.5 – 71.9%. This performance level is consistent with participants' judgment accuracy in an AGL study with manipulations of this same type of grammar when participants are tested with ungrammatical items containing few rule violations (Jamieson & Mewhort, 2009). In post-study questioning, only four participants disclosed that they had noticed any general pattern in the sequence but were not able to verbalize even a single instance of an adjacency pair (or full string), suggesting that their performance at test was not the product of explicit recall or well-formulated meta-knowledge. Next, we use scores on this adjacency pair test to assess whether and how variation in adjacency learning may be associated with differences in processing non-local linguistic dependencies.

Experiment 2: Relationship between Adjacent Statistical Learning and Language Processing

Sensitivity to long-distance relationships is indispensable to natural language processing. To determine how adjacency learning relates to the processing of nonadjacencies in natural language, we chose two sentence types involving

long-distance dependencies with (1) and without (2) potentially conflicting adjacency information:

1. a) The key to the cabinets was rusty from many years of disuse. (Mismatch)
b) The key to the cabinet was rusty from many years of disuse. (Match)
2. a) The reporter that attacked the senator admitted the error. (SR)
b) The reporter that the senator attacked admitted the error. (OR)

As illustrated by (1), a number-marked subject (*key*) in an English sentence is required to agree with the number-marking of its verb (*was*) irrespective of the numerical marking of any intervening material (e.g., *to the cabinet/s*). Although such agreement processes are pervasive in regular speech (occurring once every 5 seconds or less; Eberhard, Cutting, & Bock, 2005), production errors in agreement can be elicited for constructions such as (1a) (e.g., as in Bock & Miller, 1991). Similarly in sentence comprehension, when a sentence's head noun is singular, individuals take longer to read the verb in a "Mismatch" condition (1a) where the 'distracting' local noun (*cabinets*) mismatches in number (i.e., is plural) than in a "Match" condition (1b) where the local noun (*cabinet*) matches the head noun's number (i.e., is singular) (e.g., Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey & Bock, 1999). Thus, long-distance dependencies are created by interposing a prepositional phrase (e.g., *to the cabinet/s*) with potentially conflicting adjacency information (in the Mismatch condition) between the subject-noun and its respective verb.

In the second sentence type, long-distance dependencies are created by the center-embedding of subject-relative (SR; 2a) and object-relative (OR; 2b)

clauses that differ with respect to the manner in which the embedded verb *attacked* relates to its object. This involves a more complex, backwards-tracking long-distance dependency (to the head-noun) for ORs. In prior studies using materials resembling those in (2a-b), greater processing difficulty was elicited at the main verb of ORs compared to that of SRs, with considerable individual differences in the magnitude of this effect (e.g., King & Just, 1991; Wells, Christiansen, Race, Acheson & MacDonald, 2009).

Statistical learning has been explicitly proposed to underlie phenomena associated with both subject-verb agreement production (Haskell, Thornton, & MacDonald, 2010) and subject/ object relative-clause comprehension (e.g., Wells et al., 2009). But how might individual differences in statistical learning of adjacent statistics be connected to differences in the processing of these long-distance dependencies? The aim of Experiment 2 is to investigate whether and how differences in predictive adjacency learning are empirically related to on-line processing of these natural language contexts.

Method

Participants

The same participants from Experiment 1 also completed a subsequent natural language task. Data was omitted for participants previously excluded in Experiment 1 and from three others (2 for bilingual status and 1 for declining further participation).

Materials

There were four sentence lists, each consisting of 9 practice items, 40 experimental items, and 70 filler items. The experimental items were sentences drawn from previous studies of sentence processing: 20 subject-object relative clauses (SOR; Wells et al., 2009) and 20 subject-verb agreement transitives (S-V; Pearlmutter et al., 1999). A yes/no comprehension probe followed each item. Item conditions within sentence sets were counterbalanced across lists.

Procedure

Each participant was randomly assigned to a list, whose items were presented in random order using a standard word-by-word, moving window, self-paced reading paradigm (Just, Carpenter, & Woolley, 1982). Millisecond reading times (RTs) per word and comprehension accuracy were recorded for analyses.

Results and Discussion

Overall comprehension accuracy across participants was high, $M = 87.4\%$, $SD = 7.6$. After removing RTs in excess of 2500 ms (0.2% of data), RTs were length-adjusted for the number of characters in a word by computing a regression equation for each participant and subtracting observed RT values from predicted values (Ferreira & Clifton, 1986). Length-adjusted RTs were then examined for the same sentence regions as those in Wells et al. (2009) and Pearlmutter et al. (1999). RTs connected with relevant effects for each sentence

set were subsequently used to probe for associations with individuals' adjacency learning scores from Experiment 1.

Subject-Verb Agreement

RTs across words in the Match and Mismatch conditions, as shown in Figure 3.3, were comparable to those in Pearlmutter et al. (1999). A 34 ms effect of match (i.e., the RT difference from Mismatch minus Match conditions) was obtained at the verb ("*was*"), $F(1, 21) = 31.28, p < .0001$, which replicated Pearlmutter et al.'s findings. Strikingly, there was a significant correlation between adjacency learning scores and the Match/Mismatch RT difference at the verb ($r = .51, p = .02$), with better adjacency learning corresponding to a larger difference (see Figure 3.4). No speed-accuracy tradeoff was apparent, as comprehension accuracy for the S-V agreement set and critical RTs related to the Match/Mismatch effect were uncorrelated (all p 's $> .22$). There was also no relationship between adjacency scores and comprehension accuracy for any of the sentence sets (S-V: $r = -.02, p = .94$; SOR: $r = -.32, p = .15$; and filler items: $r = -.21, p = .35$), suggesting that the relationship between adjacency performance and reading times cannot be accounted for by offline comprehension differences.

Subject-Object Relatives

Group RTs for analyzed regions of the SOR clauses reproduced the overall patterns previously reported in the literature, as illustrated in Figure 3.5. Results replicated the main effect for clause-type at the main verb ("*admitted*")

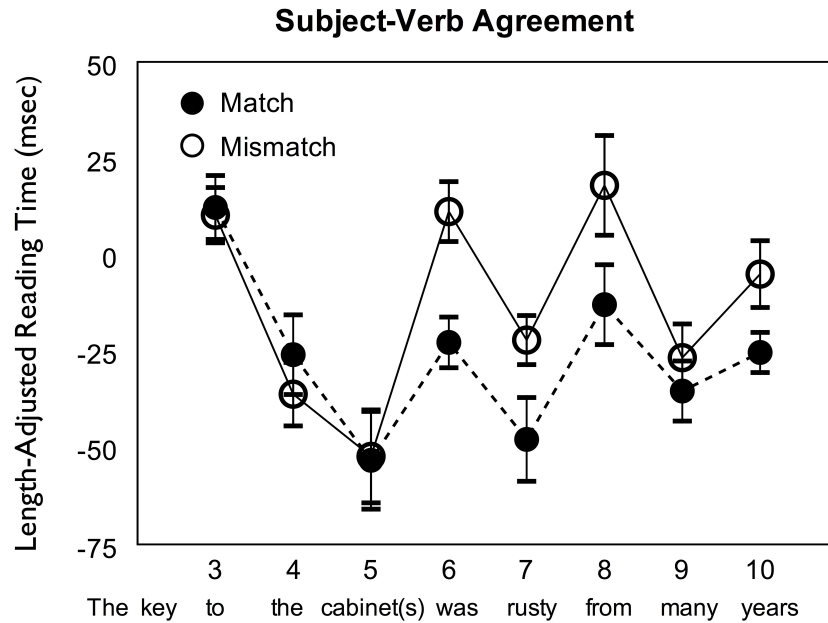


Figure 3.3. Reading times by word position for each condition (match/mismatch) of the subject-verb agreement sentences. Error bars indicate standard errors.

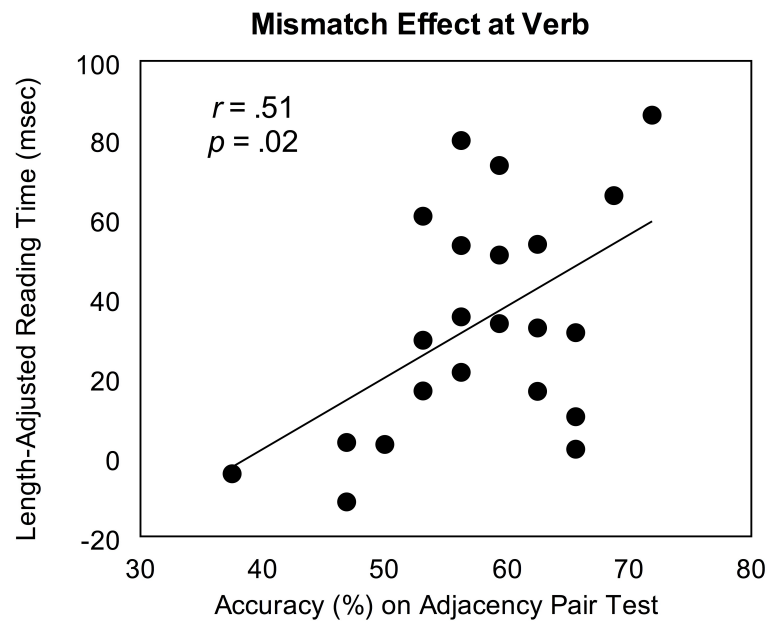


Figure 3.4. Correlation between adjacent statistical learning scores and reading times for the effect of mismatch at the verb (within the subject-verb agreement set).

from Wells et al. (2009), $F(1, 21) = 5.55$, $p = .03$, with higher RTs (by 91 ms) for ORs than SRs. However, there was no significant correlation between adjacency learning scores and main verb RTs for either SR ($r = .04$, $p = .85$) or OR ($r = -.16$, $p = .47$) sentences, nor between adjacency scores and each participant's average difference in main verb RTs between SR and OR conditions ($r = -.20$, $p = .38$). There was no evidence of a speed-accuracy tradeoff within the set, as all correlations between comprehension accuracy for subject-object relatives and the dependent RT measures were nonsignificant (all p 's $> .29$). Thus, there was no evidence associating variation in statistical adjacency learning with variation in long-distance dependency processing in the absence of potentially conflicting adjacency information.

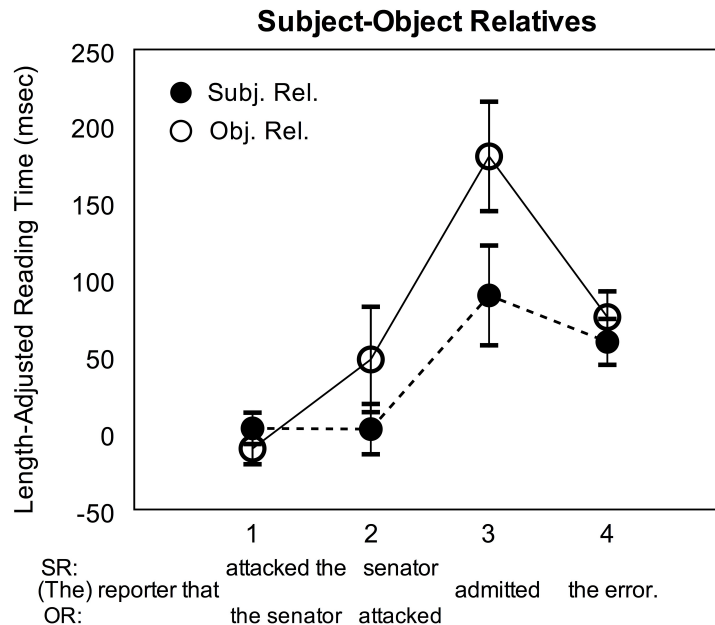


Figure 3.5. Reading times for the sentence regions of subject- (SR) and object- (OR) relative clauses. Error bars indicate standard errors.

Together, our results suggest that while good adjacency learners do not differ from less proficient learners on processing long-distance relations as such, their increased sensitivity to local relations appears to interfere with the processing of the longer-distance elements within the sentence under conflicting conditions. The irrelevant number marking of an intervening noun adjacent to the verb negatively affects the better adjacency learners' processing of the long-distance dependency between the subject noun and the verb. This effect of mismatch reflects the phenomenon of number "attraction" induced by the local noun in the intervening clause (Bock & Miller, 1991). Since the local noun is irrelevant for tracking subject-verb agreement, larger effects are indicative of less efficient language processing. The continuous correlation between adjacency scores and the critical RT measure of the S-V set thus suggests that the more sensitive a learner is to adjacent statistics, the more susceptible that individual is to interference effects from locally mismatched nouns, resulting in poorer processing of the long-distance dependency between the initial noun and verb. Crucially, however, this outcome results from enhanced sensitivity to adjacent patterns across both domains. That is, sensitivity to statistical *adjacencies* transfers to sensitivity for *local* relations in natural language – but comes at the cost of increased susceptibility to local information that may interfere with *nonadjacency/long-distance* processing.

General Discussion

While many statistical learning tasks document robust learning of adjacencies, it is unknown how such abilities directly transfer into everyday language processing. We used a within-subjects design to investigate individual differences in the statistical learning of adjacent predictive dependencies and their relationship to processing natural language dependencies. By contrasting the processing of long-distance dependencies in sentences with and without potentially conflicting adjacency information, we found that while adjacency learning skills may not be directly connected to long-distance processing per se, adjacency learning is inversely related to processing efficiency in contexts where local information may lead one astray. That is, better adjacency learners showed no difference from poorer learners processing long-distance dependencies interspersed by relative clauses, but were substantially slowed in their processing of non-local subject-verb agreement when mismatched local items intervened.

As our study design is correlational, the possibility that another factor may account for the main findings cannot be ruled out in principle. However, the *inverse* relationship between adjacent statistical learning and subject-verb agreement processing is counterintuitive. Crucially, our findings cannot readily be explained by differences in working memory, motivation, alertness, or overall language ability because better adjacency learners performed more poorly on the Match/Mismatch stimuli and did not differ in either comprehension accuracy or in on-line performance on the subject-object relative clause material. Thus, proficient adjacency learners do not have any

evident problem with tracking non-adjacent structure per se. Rather, they appear to focus too much on tracking adjacent dependencies, interfering with efficient resolution of non-local relationships when nearby (i.e., adjacent) distractive elements are present. Even though our correlational design does not allow us to determine the direction of causation, this “adjacency interference” explanation seems more plausible compared to the alternative suggestion that less efficient language processing should result in better statistical learning.

Furthermore, our results converge with recent findings indicating that certain biases in statistical learning may map onto language processing in specific ways. In a prior study, sensitivity to *nonadjacent* statistical patterns correlated with better processing of *nonadjacent* dependencies in natural language (Misyak, Christiansen, & Tomblin, 2010). Analogously, in the current study, *adjacent* statistical learners showed enhanced sensitivity only to *adjacent*, but not nonadjacent, patterns in natural language. Thus, the combined findings suggest that individual differences in processing biases for the integration of competing constraints among adjacent and nonadjacent dependencies may contribute to variation across statistical learning-linked language processing skills. As such, they speak to an open issue regarding whether different systems or processing biases may be entailed by adjacent and nonadjacent processing capabilities in humans. It has been proposed, for instance, that the two forms of processing may be subserved by separate brain areas (Friederici et al., 2006), or that the two types of statistical learning are only nominally distinct as the outcome of task-specific attention processes that

selectively hone in on adjacent or nonadjacent statistics (cf. Pacton & Perruchet, 2008). The pattern of association between adjacency learning and aspects of language processing in the current study suggests that future individual differences research incorporating careful attention to a diversity of natural dependency-structures may be needed to help establish the proper relation between adjacent and nonadjacent manifestations of statistical learning and the extent to which they may ‘tap’ into the same underlying mechanisms.

More generally, these findings may have relevance to non-language domains. Statistical learning has been demonstrated in a diverse array of contexts, from tactile sequence learning, to musical tone processing and the segmentation of human action sequences (Baldwin, Andersson, Saffran, & Meyer, 2008; Conway & Christiansen, 2005; Creel, Newport, & Aslin, 2004). Given that the sequential structure of language has parallels to the organization of temporal sequences in many other behavioral domains (Lashley, 1951), differences in sequential learning skills and the corresponding tracking of language structure may be suggestive of performances in nonlinguistic statistical learning contexts as well. Moreover, long-distance dependencies are found within the structure of complex activities such as event planning and means-ends analysis (Gómez, 2002). Our findings might therefore suggest that individuals who are good at learning adjacent sequences may be especially susceptible to local distractor effects when performing everyday activities involving longer-range or nonadjacent sequential contingencies. Future research is needed to determine the extent to

which individual differences in adjacent and nonadjacent statistical learning may be predictive of differential performances in a wide array of cognitive tasks beyond language.

REFERENCES

- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382-1407.
- Bialystok, E., Craik, F. I. M., Klein, R. & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19, 290-303.
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45-93.
- Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, 121, 127-132.
- Conway, C. M., Bauernschmidt, A., Huang, S. S. & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114, 356-371.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24-39.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119-1130.
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112, 531-559.

- Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Fiser, J. & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458-467.
- Friederici, A. D., Bahlmann, J., Heim, S., Schibotz, R. I. & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, 103, 2458-2463.
- Gebhart, A. L., Newport, E. L. & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16, 486-490.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, 114, 151-164.
- Jamieson, R. K. & Mewhort, D. J. K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 9-23.
- Jamieson, R. K. & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, 62, 550-575.

- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116, 321-340.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48, 171-184.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.
- Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35-B42.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19, 1247-1252.
- Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31, 481-507.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-131). New York: Wiley.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302-331. doi: 10.1111/j.1467-9922.2010.00626.x

- Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138-153.
- Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, 36, 569-587.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Pacton, S. & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 80-96.
- Pearlmutter, N. J., Garnsey, S. M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427-456.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233-238.
- Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2010). The effects of mood, cognitive style, and cognitive ability on implicit learning. *Learning and Individual Differences*, 20, 215-219.
- Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.
- Saffran, J. R., Johnson, E. K., Aslin, R. N. & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70,

27-52.

- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A. & Barrueco, S. (1997).
Incidental language learning: Listening (and learning) out of the corner
of your ear. *Psychological Science*, 8, 101-105.
- Shafro, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual
sequence learning in infancy: Domain-general and domain-specific
associations with language. *Infancy*, 17, 247–271.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local
syntactic coherence on sentence processing. *Journal of Memory and
Language*, 50, 355-370.
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huottilainen, M. (2009).
Statistical language learning in neonates revealed by event-related
brain potentials. *BMC Neuroscience*, 10, 21.
- Thomas, K. M. & Nelson, C. A. (2001). Serial reaction time learning in
preschool- and school-age children. *Journal of Experimental Child
Psychology*, 79, 364-387.
- Thompson, S. P. & Newport, E. L. (2007). Statistical learning of syntax: The
role of transitional probability. *Language Learning and Development*, 3, 1-
42.
- Treccani, B., Argyri, E., Sorace, A. & Della Sala, S. (2009). Spatial negative
priming in bilingualism. *Psychonomic Bulletin & Review*, 16, 320-327.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J. & MacDonald, M.
C. (2009). Experience and sentence processing: Statistical learning and
relative clause comprehension. *Cognitive Psychology*, 58, 250-271.

CHAPTER 4

Sequential Expectations: The Role of Prediction-Based Learning in Language⁸

Most individuals can relate to the common, albeit occasionally vexing, experience of having someone else anticipate and finish one's own sentence before one has completed saying it. Such behavior is but one simple reflection of the human “drive to predict,” which may serve as a “powerful engine for learning and provides important clues to latent abstract structure” (Elman, 2009, p. 572). The broader processes underlying such ordinary acts have accordingly received attention as an integral component for successful learning, understanding and use of language. For example, implicit learning of sequential regularities has been linked to an individual's ability to utilize contextual and lexically predictive information in comprehending spoken language; listeners who are better at extracting statistical relationships contained within an aural sequence are also more adept in predicting the sentence-final words of a noisy speech signal (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Across other areas of language, empirical data suggest that learned knowledge of probabilistic structure forms the basis for generating implicit expectations of upcoming linguistic input, and that the on-line engagement of such predictive skills comprises an important role in

⁸ Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138-153. Copyright © 2009 by Cognitive Science Society, Inc. Adapted with permission. As this chapter was adapted from the journal-approved manuscript, there may be minor deviations from the final copy-edited version; the published journal version should be consulted for verbatim quotes.

language acquisition and processing (for reviews, see Federmeier, 2007; Kamide, 2008; Van Berkum, 2008).

Statistical learning mechanisms that have been proposed for tracking predictive dependencies in language (Saffran, 2001; for reviews, see Gómez & Gerken, 2000; Saffran, 2003) may thus be viewed as tapping into this prediction-based process. More generally, outside of language, sequence-learning work has similarly examined basic abilities for the rapid anticipation of discrete, temporal elements under incidental learning conditions. However, while traditional artificial grammar learning (AGL; Reber, 1967) tasks have been fruitfully deployed towards studying statistical learning, they fail to provide a clear window onto the temporal dynamics of the learning process. In contrast, serial reaction time (SRT; Nissen & Bullemer, 1987) tasks have been used widely in sequence-learning research to trace individuals' trial-by-trial progress, but primarily with a focus on learning fixed, repeated structure. Despite their natural commonalities then, rarely have methodological advantages of each paradigm been jointly subsumed under a single task for exploring the on-line development of prediction-based learning.

Nonetheless, notable exceptions include the work of Cleeremans and McClelland (1991), who implemented a noisy finite-state grammar within a visual SRT task to study the encoding of contingencies varying in temporal distance; and of Hunt and Aslin (2001), who employed a visual SRT paradigm for examining learners' processing of sequential transitions varying in conditional and joint probabilities. Moreover, Howard, Howard, Dennis and Kelly (2008) adapted the visual SRT task to manipulate the types of statistics

governing triplet structures; and Remillard (2008) controlled nth-order adjacent and nonadjacent conditional information to probe SRT learning for visuospatial targets. Across these studies, participants evinced complex, procedural knowledge of the sequence-embedded relations upon extensive training over 20, 48, 6 or 4 sessions, respectively, spanning separate days. Reaction time measures collected throughout exposure enabled insights into the processing of the predictive dependencies.

In similar vein, we employ a novel paradigm that directly implements an *artificial language* within a two-choice SRT task. Distinct from previous statistical learning methods, our paradigm specifically aims to reveal the continuous timecourse of statistical processing, rather than contrasting or altering the types of statistics. The paradigm is designed for the *brief* exposure periods typical of many AGL studies and flexibly accommodates the use of *linguistic* stimuli-tokens and *auditory* cues. More generally, the task shares similarities to standard AGL designs in the language-like nature of string-sequences, the smaller number of training exemplars, and the greater transparency to natural language structure. Crucially however, it uses the dependent variable of reaction times and an adapted SRT layout to indirectly assess learning while focusing attention through a cover task. By coupling strengths intrinsic to AGL and SRT methods respectively, the “AGL-SRT paradigm” is intended to complement existing approaches to research on the statistical learning of predictive relations.

Understanding how learners process nonadjacent dependencies constitutes an ongoing area of such work, with importance for theories

implicating statistical learning in language. Natural language characteristically contains many long-distance dependencies that proficient learners need to track on-line (e.g., subject-verb agreement, embedded clauses, and relations between auxiliaries and inflectional morphemes). Even with the growing bulk of statistical learning work aiming to address the acquisition of nonadjacencies (e.g., Gómez, 2002; Newport & Aslin, 2004; Onnis, Christiansen, Chater & Gómez, 2003; Pacton & Perruchet, 2008; *inter alia*), it is yet unknown exactly how such learning unfolds, the precise mechanisms subserving it, and the degree to which statistical learning of nonadjacencies empirically relates to natural language processing.

Our AGL-SRT paradigm offers a novel entry point into the study of statistical nonadjacency learning by augmenting current knowledge with finer-grained, temporal data to illuminate how nonadjacent dependencies may be processed and anticipated over time. As such, Experiment 1 studies the timecourse of nonadjacency learning, using our novel AGL-SRT paradigm and incorporating a “prediction task” (rather than the kind of standard grammaticality test typically used; e.g., Gómez, 2002). Subsequently, Experiment 2 shows how the prediction-based, associative learning principles exemplified by simple recurrent networks closely accommodate human performances on this prediction task. Experiment 3 then probes the relevance of statistical prediction-task performance to on-line natural language processing.

Experiment 1: Statistical Learning of Nonadjacencies in the AGL-SRT Paradigm

In infants and adults, it has been established that relatively high variability in the set-size from which an “intervening” middle element of a string is drawn facilitates learning of the nonadjacent relationship between the two flanking elements (Gómez, 2002). That is, when aurally familiarized to artificial strings of the form aXd and bXe , individuals show sensitivity to the nonadjacencies (i.e., the a_d and b_e dependencies) when the set of elements from which X is drawn comprise a large set of exemplars (e.g., $|X| = 18$ or 24). Performance is poorer, however, when variability of the set-size for the X is intermediate (e.g., $|X| = 12$) or low (e.g., $|X| = 2$). Similar facilitation in high-variability conditions have also been documented for adults when the grammar is alternatively instantiated with visual shapes as elements (Onnis et al., 2003). Thus, although past research has begun to document learning in specific contexts for both infants and adults, we know little about the timecourse for acquiring predictive nonadjacencies as it actually unfolds. Here, we employ our novel AGL-SRT paradigm towards that aim.

Method

Participants

Thirty monolingual, native English speakers from the Cornell undergraduate population (age: $M = 20.6$, $SD = 4.2$) were recruited for course credit.

Materials

Throughout training, participants observed auditory-visual strings (composed of three nonwords) belonging to the artificial high-variability, nonadjacency language of Gómez's (2002). Strings therefore had the form aXd , bXe , and cXf , with ending nonword-items (d , e , f) predictably dependent upon beginning nonword-items (a , b , c). Monosyllabic nonwords (*pel*, *dak*, *vot*, *rud*, *jic*, and *tood*) instantiated the string-initial and final stimulus tokens (a , b , c ; d , e , f); bisyllabic nonwords (*wadim*, *kicey*, *puser*, *fengle*, *coomo*, *loga*, *gople*, *taspu*, *hifam*, *deecha*, *vamey*, *skiger*, *benez*, *gensim*, *feenam*, *laeljeen*, *chila*, *roosa*, *plizet*, *balip*, *malsig*, *suleb*, *nilbo*, and *wiffle*) instantiated the set of 24 middle X-tokens. The assignment of particular tokens (e.g., *pel*) to specific stimulus variables (e.g., the c in cXf) was randomized across participants to avoid learning biases attributable to the specific sound properties of words. Auditory forms of the nonwords were recorded by a female native English speaker with equal lexical stress and length-edited to 500 and 600 msec for mono- and bi-syllabic nonwords respectively. Written forms of nonwords were presented in Arial font (all caps) with standard spelling, and appeared on a computer screen that was partitioned into a 2x3 grid of uniform rectangles, as depicted in Figure 4.1. The leftmost column of the computer grid contained only the initial items of strings (a , b , c), the center column the middle items ($X_1 \dots X_{24}$), and the rightmost column the final items (d , e , f). Ungrammatical strings were generated by substituting an incorrect final element that disrupted the nonadjacency relationship, thus producing strings of the form: $*aXe$, $*aXf$, $*bXd$, $*bXf$, $*cXd$ and $*cXf$.

<u>DAK</u>	WADIM	<u>TOOD</u>
PEL	<u>FENGLE</u>	RUD

Figure 4.1. The grid display for presenting the stimulus strings on each trial. In this example, “DAK” and “PEL” are initial-string items (*a*, *b*, or *c* elements) appearing in the leftmost column; “FENGLE” and “WADIM” are middle-string items (belonging to the set of 24 X-elements) appearing in the center column; and “TOOD” and “RUD” are final-string items (*d*, *e*, or *f* elements) appearing in the rightmost column. For expository purposes only, some nonwords are underlined here to distinguish the target string (*dak fengle tood*) from the foil string (*pel wadim rud*) in this example.

Procedure

Each trial began by displaying the computer grid with a written nonword centered in each rectangle, with each column containing a nonword from a correct (target) and an incorrect (foil) stimulus string. Positions of targets and foils were randomized and counterbalanced such that they were contained equally often within the upper and lower rectangles. Only the set of items that could legally occur within a given column (initial, middle, final) were used to draw the foils. E.g., for the string *dak fengle tood*, the leftmost column might display DAK and the foil PEL, the center column FENGLE and the foil WADIM, and the rightmost column TOOD and the foil RUD, as shown in Figure 4.1.

After 250 msec of familiarization to the six written nonwords, auditory versions of the three nonwords were played over headphones. Participants used a computer mouse to click inside the rectangle containing the correct

(target) written nonword as soon as they heard it, with instructions emphasizing both speed and accuracy. The first nonword (e.g., *dak*) was played automatically after the familiarization period, whereas the subsequent two nonwords were played once the participant had responded to the previously played word (e.g., *fengle* was played after a response was recorded for *dak*, and *tood*, in turn, after the participant responded to *fengle*). Thus, when listening to *dak fengle tood*, the participant should first click DAK upon hearing *dak* (Fig. 4.2, left), then FENGLE when hearing *fengle* (Fig. 4.2, center), and finally TOOD after hearing *tood* (Fig. 4.2, right). After the participant clicks the rightmost target, the screen clears and a new set of nonwords appears 750 msec later.

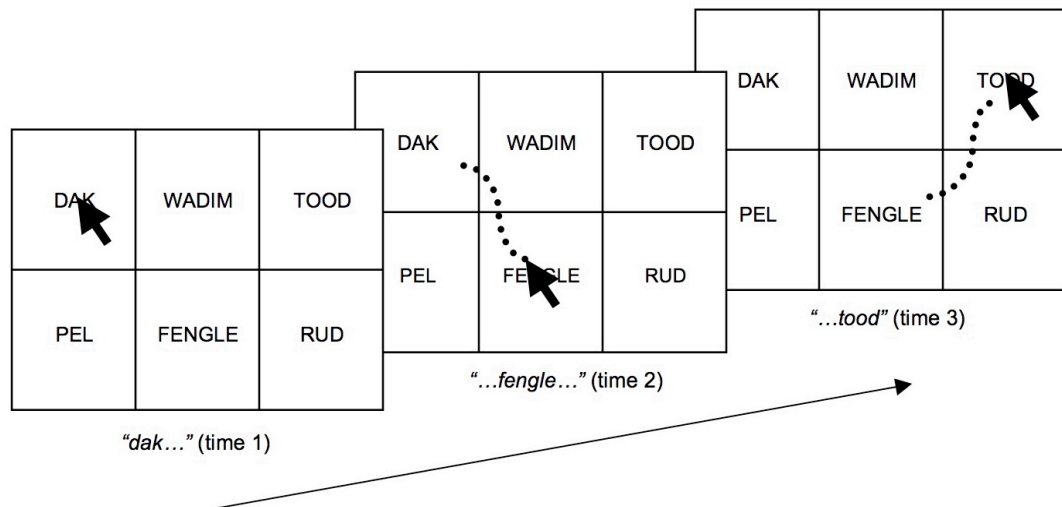


Figure 4.2. The sequence of mouse clicks associated with the auditory stimulus string "dak fengle tood" for a single trial. All trials for each of the blocks (training, ungrammatical, and recovery) followed this general pattern of sequence clicks (from left, to center, to right column clicks corresponding to the selections for the respective elements of a target string).

Per design, each nonword occurs equally often (within a column) as a target and as a foil. Thus, participants cannot anticipate beforehand which is the target and which is the foil for the first two responses of a given trial (leftmost and center columns). However, following the rationale of standard SRT experiments, if participants learn to anticipate the nonadjacent dependencies inherent in the stimulus strings, then they should respond increasingly faster to the final target. As our dependent measure, we thus recorded on each trial the reaction time (RT) for the predictive, final element, subtracted from the RT for the non-predictive, initial element to control for practice effects and serve as a baseline.

Participants were first exposed to 6 training blocks, each of which consisted of a random presentation of 72 unique strings (24 strings x 3 dependency-pairs), for exposure to a total of 432 grammatical strings. After this, participants were presented with 24 ungrammatical strings, with endings that violated the nonadjacent dependency (in the manner noted above). A final training “recovery” block of 72 grammatical strings then followed this brief ungrammatical block. Transitions between all blocks were seamless and unannounced.

Upon completing the 8 exposure blocks, participants performed the “prediction task” of key interest here because it provides a direct measure of the degree to which participants have learned the nonadjacency patterns. They were told that there were rules specifying the ordering of nonwords for the auditory sequences, and were asked to indicate the endings for 12 subsequent strings upon being cued with only the first two sequence-elements. In other

words, participants observed the same grid display as before and followed the same procedure for the non-predictive initial and middle columns (e.g., selections corresponding to *dak fengle...* in Fig. 4.2), but then had to select which nonword in the predictive final column (e.g., TOOD or RUD) they thought best completed the string without hearing the ending (and without feedback).

Results

Since instructions emphasized speed in addition to accuracy, there was a small proportion of errors made by participants, as is commonly reported in SRT studies. Thus, only accurate string trials (with only one selection response for each of the three targets) were used for analyses. These averaged 90.0% ($SD = 5.6$) of training block trials, 84.7% ($SD = 15.7$) of ungrammatical trials, and 87.1% ($SD = 12.3$) of recovery trials.⁹ Final-element RTs were subtracted from initial-element RTs on each trial, with means of these resultant RT difference scores computed for each block. Figure 4.3 plots group averages for these difference scores, with positive values reflecting nonadjacency learning.

A one-way repeated-measures analysis of variance (ANOVA) with block as the within-subjects factor was performed on mean RT difference scores. Mauchly's test indicated a violation of the sphericity assumption ($\chi^2(27) = 111.82, p < .001$), so Greenhouse-Geisser estimates ($\epsilon = .36$) were used

⁹ As analyzed trials required accuracy for all 3 string-elements composing a string-trial (rather than for single-selection responses defining one "trial" in standard SRT designs), this criterion is quite conservative, and may underestimate participants' total accuracy across all single responses. E.g., final-element selection accuracy across trial-types was 95.9% (2.4), 93.2% (6.5), and 94.2% (6.1).

to correct degrees of freedom. There was a significant effect of block on RT scores, $F(2.55, 73.96) = 8.97, p < .001$. As shown in Fig. 4.3, RT differences gradually increased across blocks, albeit with an expected performance decrement in the ungrammatical 7th block. As also evidenced by the group trajectory, sensitivity to nonadjacent dependencies required considerable exposure (an average of 5 blocks) before reliably affecting responses; this is consistent with Cleeremans and McClelland's (1991) finding that learning for long-distance contingencies emerges less rapidly than for adjacent dependencies.

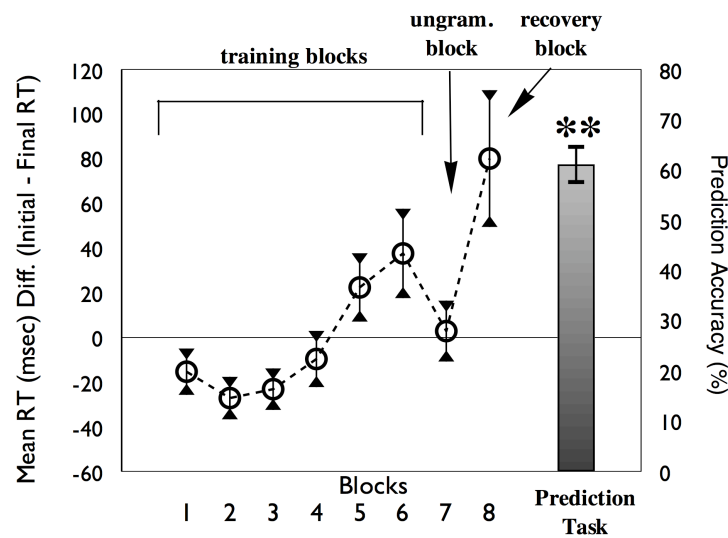


Figure 4.3. Group learning trajectory (as a plot of mean RT difference scores) and prediction accuracy in Experiment 1.

Following interpretations in the sequence learning literature for comparing RTs to structured versus unstructured material (e.g., Thomas & Nelson, 2001), we specifically assessed performance differences across the

final training block, ungrammatical block and recovery block. Planned contrasts confirmed that mean RT differences significantly decreased in the ungrammatical block compared to performances in both the preceding training block, $t(29) = 2.11$, $p = .04$, and succeeding recovery block, $t(29) = 3.22$, $p < .01$. This relative performance drop in the ungrammatical block (Block 6 minus Block 7: $M = -34.8$ ms, $SE = 16.5$; Block 8 minus Block 7: $M = 77.3$ ms, $SE = 24.0$ ms) provides a confirmation of nonadjacency learning using an established SRT measure.

Of central focus to the interrelated experiments that follow next, accuracy scores on the prediction task were calculated for each individual. Participants averaged 61.1%, with a large standard deviation (21.4%) and group range (25 - 100%) reflecting substantial interindividual variation. Group-level performance was above chance, ($t(29) = 2.85$, $p < .01$), providing a gauge of predictive skills for anticipating the statistical nonadjacencies. But what kind of computational mechanism may subserve the kind of learning evidenced by this prediction task and, more generally, by the on-line AGL-SRT task? We address this question in Experiment 2, before going on to show in Experiment 3 that the performance on the prediction task provides a sensitive index of individual differences in on-line language processing.

Experiment 2: Computational Simulations of On-line Nonadjacency Learning

The new paradigm in Experiment 1 highlights the gradual statistical learning of nonadjacencies in prediction-based performance; however, the

computational mechanisms that can accommodate such findings remain to be investigated. Cleeremans and McClelland (1991) have previously shown that the simple recurrent network (SRN; Elman, 1990) can capture performance on AGL-like SRT tasks. Furthermore, the anticipation of unfolding temporal structure and implicit prediction-based feedback are distinctive, fundamental features of the SRN's associative architecture (see, e.g., the discussion in Altmann & Mirković, 2009). We thus chose to closely model on-line performance from our task with SRN simulations based on the *exact* same exposure and input as in the human case.

The SRN is essentially a standard feed-forward network equipped with context units containing a copy of hidden unit activation at the previous timestep, thus providing partial recurrent access to prior internal states. The context layer's limited maintenance of sequential information over past timesteps allows the SRN to potentially discover temporal contingencies spanning varying distances in the input. Next, we use the SRN's graded output values and prediction-based learning mechanism to model human RTs and prediction scores from Experiment 1.

Method

Networks

Simulations were conducted with 30 individual networks, one corresponding to each human participant, and each randomly initialized with a different set of weights within the interval (-1,1) to approximate learner differences.

Localist representations were employed for the 30 input and output units,

with one unique unit corresponding to each nonword. The hidden layer had 15 units. The networks were trained using standard backpropagation with a learning rate of 0.1 and momentum at 0.8.

Materials. The SRNs received the same input as human participants, presented using the same randomization process as in Experiment 1, and tested on the same “prediction task” strings (with the same target-foil pairings).

Procedure. Networks received the exact same amount of exposure to the statistical dependencies as the human participants (i.e., 6 grammatical blocks of 72 string-trials, an ungrammatical block of 24 trials, a recovery block of 72 trials, and a 12-item prediction task)—and no additional training. Context units were reset between string-sequences by setting values to 0.5; this simulated the screen-clear and between-trial pauses that human participants observed. Weight changes were carried out continuously throughout training, except for the prediction task items at the very end, when weights were “frozen” (reflecting the fact that human participants received no auditory input/feedback for selecting the final elements of prediction-task strings).

Results

The networks’ continuous outputs were recorded, and performance was evaluated by computing a Luce ratio difference score for string-final predictions on each trial. A Luce ratio is calculated by dividing a given output-unit’s activation value by the sum of the activation values of all output

units. During processing, the representation formed at the output layer of the SRN approximates a probability distribution for the network's prediction of the next element. Thus, on the timestep where a middle (X) element is received as input, if the network has become sensitive to the nonadjacent dependencies, it should most strongly activate the output unit corresponding to the correct, upcoming string-final nonword. The Luce ratio essentially quantifies the proportion of total activity owned by this output unit.

To approximate human RT difference scores, we subtracted the Luce ratio for the foil unit from the Luce ratio for the target unit. Since networks cannot erroneously select a foil in the same way that humans occasionally do (and which were excluded from analyses, as noted earlier, in line with standard SRT protocol), accurate trials for the networks were defined as those in which the Luce ratio for the target exceeded that for the foil. As in Experiment 1, only responses/outputs from accurate trials were analyzed.

A one-way repeated-measures ANOVA with block as the within-subjects factor was performed. As Mauchly's test indicated a violation of the sphericity assumption ($\chi^2(27) = 66.947, p < .001$), degrees of freedom were corrected using Greenhouse-Geisser estimates ($\epsilon = .60$). There was a main effect of block on mean Luce ratio difference, $F(4.21, 121.96) = 35.57, p < .001$. As in the human case, difference scores gradually increased, with a performance decrement in the 7th (ungrammatical) block. This drop was significant in relation to both the preceding and succeeding grammatical blocks, $t(29) = 6.76, p < .0001$; $t(29) = 7.80, p < .0001$.

The networks' mean Luce ratio difference scores across blocks are plotted in Figure 4.4, alongside the human learning trajectory from Experiment 1.¹⁰ Both trajectories are indicative of a gradually developing sensitivity to the nonadjacent dependencies, with a steeper ascent from blocks 4 to 6. The simulated block scores further account for 78% of the variance in human RT difference scores ($p < .01$).

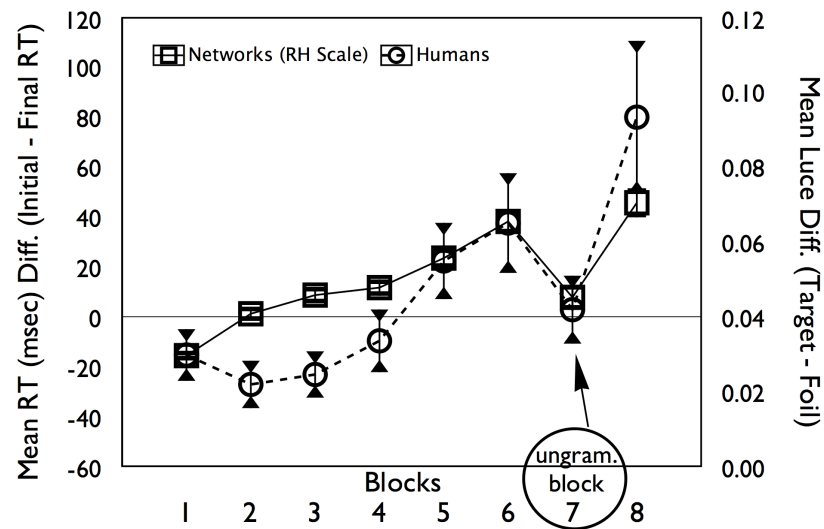


Figure 4.4. Comparison of group learning trajectories for SRN (squares) and human (circles) learners.

As the analog to the human prediction task, in which SRNs received the same test-strings with foil-pairings as the humans, we considered the network's selection to be the nonword corresponding to the unit with a higher Luce ratio (from among the 2 choices for an ending). Prediction task accuracy

¹⁰ Because the learning metric for humans subtracts final- from initial-element RTs (to control for potential motor effects) whereas that for the SRNs uses only final-element values, Y-axes are equalized with block 1 level performance as the baseline.

as a proportion correct out of the 12 items was then computed accordingly. The SRNs' scores averaged 56.4% ($SD = 13.4\%$), which was above chance-level, $t(29) = 2.61, p = .01$. As seen in Figure 4.5, the distribution of the networks' prediction scores were also not significantly different from that of humans', $t(58) = 1.025, p > .30$. Although the networks exhibited somewhat less variability, they captured the identical full range of human performance from 25 - 100% accuracy. Thus, the SRN is able to closely match human performance both across training in the AGL-SRT task as well as on the prediction task. Given that this type of connectionist model has been used extensively to model the processing of non-local dependencies in natural language (e.g., Christiansen & Chater, 1999; Christiansen & MacDonald, 2009; Elman, 1991; MacDonald & Christiansen, 2002), we next explore whether the ability to predict correct nonadjacency relations in Experiment 1 is associated with the processing of long-distance dependencies in language.

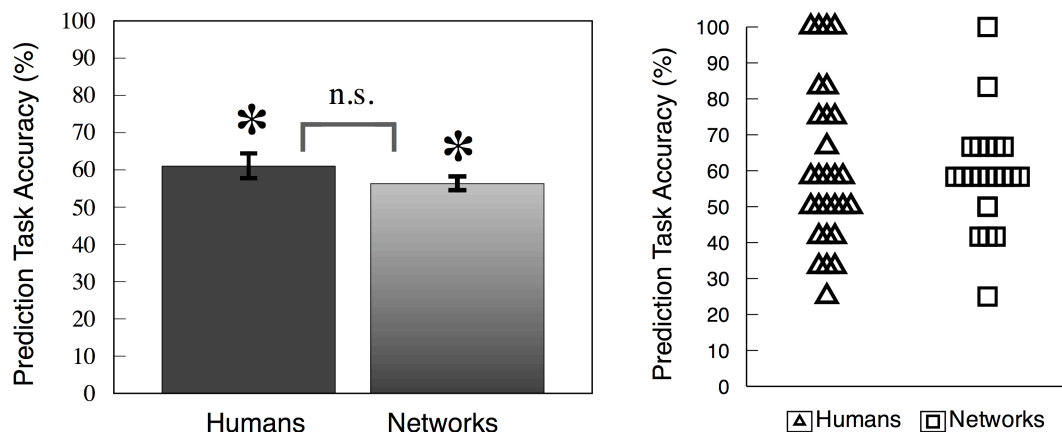


Figure 4.5. Prediction task means for humans and networks (left panel) and corresponding score distributions of both groups (right panel).

Experiment 3: Individual Differences in Language Processing and Statistical Learning

While Experiment 2 attests to the kind of computational mechanisms that may subserve performance on the AGL-SRT and prediction tasks, the relevance of the new paradigm for the processing of complex long-distance dependencies in natural language remains to be probed. In the language literature, individual differences have been prominently studied within the context of subject-object relative clause processing phenomena. Center-embedded object relative (OR) sentences (illustrated in 2) are generally more difficult to process and comprehend than subject relative sentences (SRs; such as 1), with the structural difference between the two residing in how the embedded verb (*attacked*) relates to its object (though see, Reali & Christiansen, 2007). For ORs, the embedded verb enters into a nonadjacent dependency with the nonlocal head-noun (*reporter*), whereas for SRs the embedded verb's object (*senator*) is situated more locally. The greater processing difficulty associated with ORs can be construed as a reflection of changing, probabilistic expectations for the continuation of the sentence as its temporarily indeterminate (and relative to SRs, less frequent and irregular) structure unfolds (Gennari & MacDonald, 2008).

1. The reporter *that attacked the senator* admitted the error.
2. The reporter *that the senator attacked* admitted the error.

The locus of this greater processing difficulty for ORs compared to SRs is evidenced at the main verb, where reading times (RTs) for ORs are

protracted. King and Just (1991) first reported individual differences in the degree of comparative difficulty, which they linked to verbal working memory differences on a reading span task. Interpretations of these findings, however, have been in dispute between experience-based versus capacity-based accounts (e.g., Just & Carpenter, 1992; MacDonald & Christiansen, 2002; see also Waters & Caplan, 1996).

While capacity-based views impute low-span individuals' poorer processing of ORs to limitations in memory resources, experience-based views emphasize exposure-related factors that shape linguistic expectations and modulate the processing difficulty that readers encounter. In support of the latter approach, MacDonald and Christiansen (2002) conducted SRN simulations whereby they qualitatively fit the SR/OR RT patterns attributed to low- and high-span individuals as a function of the amount of relative clause exposure received by their networks. In addition, a human training study by Wells, Christiansen, Race, Acheson and MacDonald (2009) documented that increased experience in reading relative clauses (compared to a control condition) altered participants' RT profiles towards matching those of ostensibly high-span individuals (and the aforementioned high-trained SRNs).

These studies imply a crucial role for statistical learning as a mediator of experience-driven effects on shaping readers' (probabilistic) expectations, thus facilitating subsequent RTs for ORs. If implicit prediction-based processes, as tapped by statistical learning mechanisms, are indeed important to such processing phenomena and sensitively reflected in prediction-task

scores from our AGL-SRT paradigm, then individual differences in statistical predictive skills from Experiment 1 should systematically vary with differences in relative clause processing. Experiment 3 thus empirically tests this hypothesis using a within-subjects design.

Method

Participants

The last 20 participants in Experiment 1 were recruited to participate afterwards in this experiment for additional credit. Data from four of these participants were excluded (one for refusal to participate, and three due to equipment malfunction).

Materials

SR/OR sentence pairs from Wells et al. (2009) were used to prepare two counterbalanced, experimental sentence lists. Each list contained 12 initial practice items, 40 experimental items (20 SRs, 20 ORs), and 48 filler items. Semantic plausibility information for subject/object nouns was controlled in the experimental sentences, with comprehension questions (Yes/No format) following each sentence item.

Procedure

Participants were randomly assigned to an experimental list, which was presented using a standard self-paced reading, moving-window paradigm (Just, Carpenter & Woolley, 1982). Sentence items were thus presented in

random order, with both millisecond RTs for each word and accuracy for each comprehension probe recorded.

Results

Raw RTs corresponding to practice items and those in excess of 2500 ms (1.01% of data) were excluded from analyses. RTs were length-adjusted by computing a regression equation for each participant based on the character-length of a word, and subtracting observed RT values from predicted values (Ferreira & Clifton, 1986). Means from these residual RTs were then calculated across subject- and object- relative clauses for the same sentence regions that have been analyzed in prior related work (see, e.g., Wells et al., 2009).

Consistent with past studies, greater processing difficulty for ORs was reflected by substantially increased RTs at the main verb. Also in-line with prior findings, overall comprehension rate was high (86.8%, $SD = 8.1$), with significantly poorer accuracy observed for ORs (74.7%, $SD = 19.0$) than for SRs (85.6%, $SD = 9.6$), $t(15) = 2.66$, $p = .02$.

To test our hypothesis about the involvement of statistical predictive skills in relative clause processing, we correlated individuals' prediction task scores from Experiment 1 with their length-adjusted RTs at the main verb of the relative clauses, with results illustrated in Figure 4.6. For SRs, there was no significant association ($r = -.10$, $p = .72$), as expected, because experience has not been shown to be a factor for further facilitating processing of this comparatively easier clause-type. For ORs, however, higher prediction task scores were associated with lesser reading difficulty ($r = -.59$, $p = .02$).

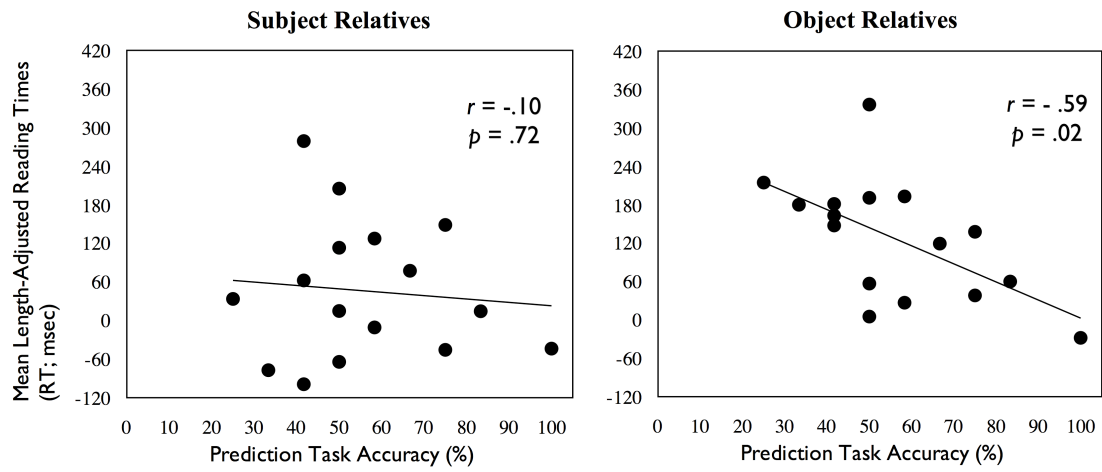


Figure 4.6. Length-adjusted reading times at the main verb of subject- (left) and object-relatives (right), plotted against prediction task scores.

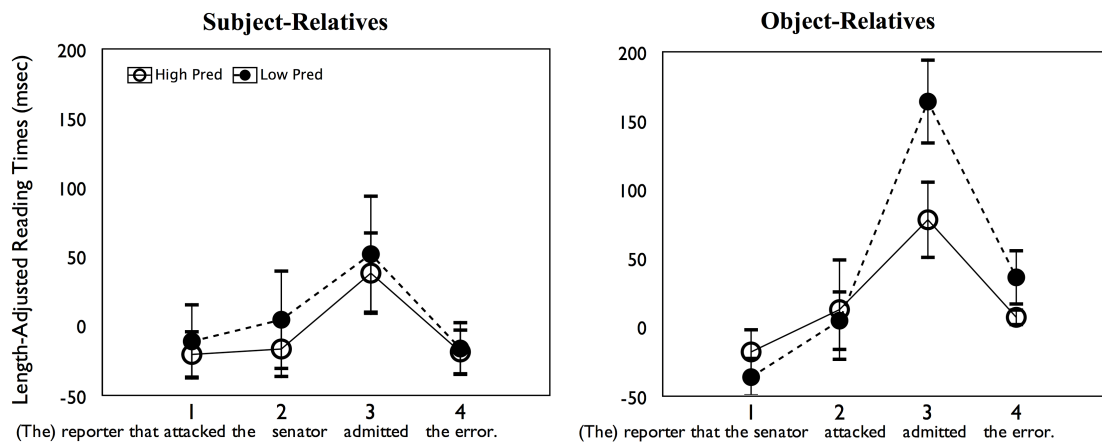


Figure 4.7. Length-adjusted reading times across sentence regions of subject-relatives (SR; left panel) and object-relatives (OR; right panel) in Experiment 3 for participants with “low” (filled circles) and “high” (open circles) prediction task scores from Experiment 1.

Moreover, individual differences in prediction task scores were not predictive of RTs for any other standard SR/OR sentence regions *except*, crucially, at the main verb of ORs—the anticipated locus of observed processing difficulty. This pattern is additionally evidenced and clearly reflected in the RTs of participants when subdivided into “high” and “low” groups based on prediction task scores (with chance-level performance of 50% as the cutoff-level). As seen in Figure 4.7, “low pred” participants ($n = 9$, $M = 42.6\%$, $SD = 8.8$) differed from “high pred” participants ($n = 7$, $M = 73.8\%$, $SD = 4.8$) only for processing at the critical OR main verb. The overall pattern of RTs for both SR and OR sentences closely mirrors qualitatively the pattern of “high” versus “low” experience participants in Well et al. (2009).

These findings support the hypothesis that prediction-based processes tapped by statistical learning mechanisms (as assessed through prediction-task performances in the AGL-SRT paradigm) are substantially involved in individuals’ on-line natural language processing. This conclusion is also corroborated by results from an individual-differences study by Misyak and Christiansen (2007), in which both adjacent and nonadjacent statistical learning performance was an even better predictor of sentence comprehension than verbal working memory span scores. The current study thus expands on those findings by documenting that differences in nonadjacent statistical learning vary systematically with the on-line tracking of nonadjacent dependencies exemplified by OR sentences.

Discussion

Nonadjacent dependency learning was investigated here across three interconnected experiments, using results from a novel AGL-SRT paradigm. The new task investigated individuals' learning of nonadjacencies as it unfolded on-line. Individual differences in performance on the statistical prediction task were shown to correlate with the processing of complex, long-distance dependencies occurring in natural language, as well as to compellingly appear to recruit upon the kind of associative-based learning principles exemplified by SRNs.

But how does the individual variation in statistical learning manifest itself in our AGL-SRT statistical learning task? Inspection of micro-level trajectories from Experiment 1 for good and poor statistical learners (as measured by prediction task scores) indicates distinct differences during nonadjacency learning.¹¹ Thus, there are contrasts in the shape of the statistical learning trajectory, final training performance, and the response to ungrammatical items. In particular, the poor prediction-task performers do not show evidence of learning until the very end of training, contributing to the strong recovery effect on this block observable in Figure 4.3. We expect that future work into such individual differences in statistical learning will benefit from closer attention to predictive processing as it unfolds over time, investigated using on-line methods such as the AGL-SRT task used here.

In broader theoretical terms, our close modeling of human performance

¹¹ See the Appendix for graphs of the trajectories corresponding to the subgroups of good and poor statistical learners.

with SRNs in Experiment 2 argues against the assumption that verbal working memory capacity operates as a basic constraint for the human results in Experiments 1 and 3; it also establishes a connection with the results from MacDonald and Christiansen (2002) in terms of common mechanisms. Their simulations with SRNs predicted that increased exposure to relative clause sentences should differentially affect ORs. Wells et al. (2009) empirically confirmed those predictions and further hypothesized that statistical learning may be centrally involved—but did not otherwise speak to what the underlying mechanisms may be. The combination of results from the three experiments reported here, however, directly supports Wells et al.'s hypothesis. In particular, not only did individual differences in statistical prediction performance correlate uniquely with on-line language processing measures at the key main verb region in OR sentences, as would be expected on an experience-based account, but prediction performance for high- and low-performing individuals on SR/OR processing also closely conformed to the pattern obtained for participants measured to have high/low verbal working memory spans in King and Just (1991), as well as those of the high/low experience manipulations for SRNs and humans in MacDonald and Christiansen and Wells et al., respectively. Together with previous findings that statistical learning overall is a better predictor of sentence processing skills than verbal working memory (Misyak & Christiansen, 2007), these results provide converging evidence for statistical learning as a key contributing factor to individual differences in language, and as a mechanism for producing sequential expectations for upcoming linguistic material.

REFERENCES

- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583-609.
- Christiansen, M. H. & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157-205.
- Christiansen, M. H. & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, 59:Suppl. 1, 129–164.
- Cleeremans, A. & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235-253.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114, 356-371.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Elman, J. L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-225.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33, 547-582.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491-505.
- Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.

- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language*, 58, 161-187.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Gómez, R. L. & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178-186.
- Howard, J. H. Jr, Howard, D. V., Dennis, N. A. & Kelly, A. J. (2008). Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1139-1157.
- Hunt, R. H. & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658-680.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Just, M. A., Carpenter, P. A. & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2/4, 647-670.

- King, J. & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.
- MacDonald, M. C. & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35-54.
- Misyak, J. B. & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1307-1312). Austin, TX: Cognitive Science Society.
- Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Onnis, L., Christiansen, M. H., Chater, N. & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pacton, S. & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 80-96.

- Real, F. & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1-23.
- Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology*, 61, 400-424.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110-114.
- Thomas, K. M., & Nelson, C. A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology*, 79, 364-387.
- Van Berkum, J. J. A. (2008). Understanding sentences in context: What brain waves can tell us. *Current Directions in Psychological Science*, 17, 376-380.
- Waters, G. S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49, 51-79.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J. & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250-271.

CHAPTER 5

Summary and Discussion

Across three interrelated studies, the empirical relationship between statistical learning and language was investigated using an individual differences approach. College-aged adults in the first, small-scale individual-differences study, described in Chapter 2, completed a battery of tasks assessing statistical learning (for adjacent and nonadjacent dependencies), language comprehension (for three different types of sentences), memory-related factors (verbal working memory and short-term memory), language-relevant variables (lexical knowledge and print exposure) and nonverbal aptitudes (fluid intelligence and cognitive motivation). Contrary to standard assumptions of intraindividual invariance in incidental learning abilities, individual differences in statistical learning were documented within the normal adult population. These differences systematically varied with a subset of the other study measures, including most strongly with those for language comprehension and verbal working memory.

While performances on the separate adjacent and nonadjacent statistical learning tasks themselves were uncorrelated in the study, individuals' sensitivity to each of these kinds of statistical regularities (i.e., adjacent or nonadjacent) positively corresponded to variations in comprehending different types of sentences (e.g., involving the tracking of either local or long-distance relationships) from the natural language task. Thus, after controlling for the effect of all other predictor variables in

regression analyses, adjacent statistical learning was the only remaining predictor for comprehending phonological typicality sentences; these were sentences with ambiguous noun/verb homonyms that were disambiguated locally. In a similar manner, nonadjacent statistical learning was the only remaining predictor for comprehending subject-object relative clause sentences, which involve integrating long-distance linguistic dependencies.

The correlational nature of the study's design cannot establish causality and the relatively small number of participants reduces the power to detect potential associations. However, despite these limitations, this small-scale individual-differences study provides an initial confirmation that systematic, statistical learning differences do exist and may potentially explain more variation in language than measures (e.g., verbal working memory) that have been the disproportionate focus for traditional accounts of individual differences in language performance. Thus, we documented individual differences in statistical learning of adjacent and nonadjacent dependencies, and these variations correspondingly mapped onto sensitivity for processing similar relations occurring in natural language—a theme also repeated in the studies presented across Chapters 3 and 4.

Statistical learning in the individual-differences study was assessed via canonical tasks in the literature, which have proven highly useful over the years. However, in order to better tap into the ongoing dynamics of the learning process, a newly developed AGL-SRT task was used to examine learning for adjacent and nonadjacent statistical dependencies across a pair of complementary studies. In both studies of “adjacency” and “nonadjacency”

processing, presented in Chapters 3 and 4 respectively, participants' learning trajectories in the AGL-SRT paradigm provided evidence for an overall gradually-developing sensitivity to the statistical regularities (either adjacent or nonadjacent) underlying observed string-sequences. An index for individual differences in statistical learning was also obtained from measures that isolated specific knowledge for conditional probabilistic dependencies (ordered bigrams in the adjacency study; distal dependency-pairs in the nonadjacency study). This statistical learning index was then used to test for within-subject associations to on-line language processing variations from a subsequent natural language task (using the self-paced reading paradigm).

Reading times from the natural language task were examined as indices of individuals' on-line processing difficulty for the parsing of various language structures, subject-object relatives and subject-verb agreement, and reproduced patterns widely documented in the literature. Following standard psycholinguistic interpretation, contrastive reading times at the point of integrating the main verb in subject-object relative clauses reflect the processing difficulty concomitant with tracking long-distant linguistic dependencies (as described in more detail in Chapters 3 and 4). Analogously, contrastive reading times at the main verb of certain subject-verb agreement sentences (i.e., Pearlmutter, Garnsey, & Bock, 1999) reflect processing difficulty concomitant with tracking long-distant dependencies in the presence or absence of a locally induced dependency mismatch (as described in more detail in Chapter 3).

Within our studies, individual differences in nonadjacent statistical

learning were associated with *less* processing difficulty for resolving long-distant dependencies in the subject-object relative clause materials. This is also consistent with the finding from the small-scale individual-differences study in which nonadjacent statistical learning was the only remaining predictor in regression analyses for comprehending subject-object relatives. However, no association to these language materials was observed for adjacent statistical learning. Conversely, individual differences in adjacent statistical learning were associated, in turn, with *greater* processing difficulty for resolving long-distant dependencies in the subject-verb agreement materials; this occurred under conditions when locally conflicting but irrelevant information was present. This seemingly counterintuitive finding for adjacent statistical learning was thus an indirect effect of greater sensitivity to local dependencies (between the main verb and adjacently-embedded noun), which increased susceptibility to local interference effects in processing the long-distance dependency between the head-noun and main verb. Therefore, as a bottom line, nonadjacent statistical learning was associated with more efficient on-line processing of nonadjacent language structure; adjacent statistical learning was associated with more sensitivity to adjacent language structure.

These findings mirror the general pattern reported in the small-scale individual-differences study where facility in processing artificial statistical dependencies maps onto variations in processing similar natural language relations. Crucially, however, some specificity in the mappings is evidenced. Skilled nonadjacent statistical learners were better at processing the long-distant dependencies entailed by subject-object relative clauses. Skilled

adjacent statistical learners, however, did not differ in their online processing of long-distance dependencies *per se* in the absence of any conflicting adjacency information. And only when conflicting local information was present did skilled adjacent statistical learners perform more *poorly* for integrating long-distance dependencies. While caution is generally warranted for interpreting null reports of association, the same subject-object relative clause materials were used across both studies, and both studies (using comparable sample sizes) replicated the robust effect in the literature for humans' processing of embedded subject-object relatives.

Although the influence of an unidentified third factor cannot be ruled out in principle, the specificity and contrastive pattern of correlations throughout weaken potential claims that additional, unaccounted-for variables could underlie performance variations across the statistical learning and natural language tasks. For instance, it seems unlikely that individual differences in either task-engagement or alertness are driving these patterns, as it would imply that more motivated or vigilant individuals performed better on the statistical learning tasks but performed *worse* on language processing (for the subject-verb agreement materials) in select contexts. Similarly, for the adjacency study, better working memory would not predict poorer performance for processing the long-distance subject-verb agreement dependencies (nor the lack of association across subject-object relatives).

Even though the adjacency study does not implicate working memory as the source for variation, one could still be inclined to argue that verbal working memory limitations play a role in mediating processing for the

longer-range statistical and linguistic contingencies in the nonadjacency study. However, the simple recurrent network (SRN) simulations in Chapter 4 converge with empirical data from this study and other work (see also below) to argue against a necessary role for a separate memory capacity in explaining humans' processing of such dependencies. That is, the SRN simulations were conducted under the same training conditions as the humans and with identical amounts of exposure to the statistical dependencies. Without any manipulation of network parameters that would correspond directly to memory variations, the SRNs closely captured the same online trajectory of humans' statistical learning, as well as the group average and full range of scores on the individual-differences index from the "prediction task." The modeling results, instead, are consistent with an associative and ("implicit") prediction-based mechanism that may commonly underlie processing across the statistical learning and language tasks.

Having summarized the findings across the experiments from Chapters 2-4, it may be worthwhile to revisit the three theoretical implications sketched at the beginning of this thesis.

Capacity- and experience-based accounts

In the *Introduction*, it was noted that individual differences in adult sentence processing have been debated with respect to whether they essentially derive from variations in (language-separate) cognitive capacities or experiences with language (i.e., exposure to specific dependency relationships). Results from

the nonadjacency study present a natural fit with experience-based accounts of subject-object relative clause processing phenomena (e.g., MacDonald & Christiansen, 2002; contra Just & Carpenter, 1992) . The discussion in Chapter 4 highlights in more detail the interconnections between the AGL-SRT findings and the prior literature; an extended treatment is also given elsewhere in another paper with a similar pattern of online findings with respect to nonadjacent statistical learning (Misyak, Christiansen, & Tomblin, 2010). As such, a lengthy rehash of an involved topic is avoided here, but it is worth underscoring the fact that the variations in nonadjacent statistical learners' online language processing patterns (see Figure 4.7) mirror exceptionally well the reading times patterns for the effects of experience in Wells, Christiansen, Race, Acheson, and MacDonald (2009), following from the predictions and simulations in MacDonald and Christiansen (2002). In turn, the convergence of these studies provides a compelling alternative account for language variation canonically attributed to working memory capacity. The strong interrelationships between statistical learning, language, and verbal working memory measures from the individual-differences study (Chapter 2) remain consistent with this newer account. (The aforementioned dispute is not about whether verbal working memory span measures are associated with relative-clause processing, but whether the memory span measure should be construed as denoting separate capacity-limitations or as tapping into processing skill that is part-and-parcel of the language system.) The nonadjacency study (Chapter 4) contributes further to this literature by implicating *individual differences* in statistically learning from experience as a

contributor to variation in relative clause processing.

Thus, more generally, the work in this thesis suggests that statistical learning itself may be a largely overlooked individual-differences factor that may mediate effects of experience on syntactic processing. Future work with an attention to a variety of other dependency-structures is needed to strengthen the generality of this claim. On the theoretical side, this notion provides a potential synthesis within larger discussions of the relative contributions of biological and experiential factors in language: namely, individual differences in the ability to learn from experience, by way of statistical learning, may be a substantial source for language variation. Conceivably as well, variations in statistical learning itself may have partly biological underpinnings.

Adjacent and nonadjacent dependencies

As outlined earlier, the lack of association between adjacent and nonadjacent statistical learning in the small-scale individual-differences study raises questions as to whether separate systems may be entailed in tracking and processing statistical dependencies. Before considering this idea further, however, a caveat is warranted. Failure to detect a significant correlation does not necessarily signify the absence of any actual relationship. Aside from potentially inadequate statistical power (due to the small sample size), performances across the two statistical learning tasks may be uncorrelated due to differences in AGL learning strategies or task demands. Additionally, performances in the standard nonadjacent statistical learning task from

Chapter 2 seemed susceptible to both a ceiling effect and bimodality in the distribution of learning scores. Further experimentation with a larger sample and more sensitive index of learner differences is required to draw conclusions regarding whether a substantial association (i.e., of meaningful effect size) exists between the two forms of statistical learning. Such an investigation is ongoing, using the newer AGL-SRT task, which has yielded more graded score distributions (and reaction times) with no ceiling effect. These tasks are presently employed in a larger-scale, individual-differences study (of approximately two hundred adult participants) that should yield informative findings.

Nonetheless, it is worthwhile considering some possibilities under the present findings. Thus, the data reviewed across Chapters 2-4 may seem compatible with a separate-systems account, such as postulated by Friederici and colleagues (Friederici, 2004; Friederici, Bahlmann, Heim, Schibotz, & Anwander, 2006). Their primary support comes from neuroscience findings suggesting that the processing of adjacent dependencies and (hierarchical) nonadjacent dependencies are subserved by separate brain areas—the left frontal operculum (for both types of dependencies) and Broca’s area (for hierarchical nonadjacent dependencies selectively). However, test items from Friederici et al’s (2006) artificial hierarchical grammar may have been plausibly discriminated by participants through behavioral strategies (counting, repetition) other than processing hierarchical embeddings (de Vries, Monaghan, Knecht, & Zwitserlood, 2008). Further in conflict with these claims, Broca’s area has shown activation for violations of artificial finite-state

grammars (Petersson, Forkstam, & Ingvar, 2004) and has been implicated in the impaired visual artificial grammar learning of individuals with agrammatic aphasia (Christiansen, Kelly, Shillcock, & Greenfield, 2010). Thus, claims for separate adjacent and nonadjacent learning systems might still be voiced in principle, but are unsupported by the particular functional localization alleged in Friederici et al.'s account.

Alternatively, Pacton and Perruchet (2008) view adjacent and nonadjacent statistical learning as the outcome of task-specific attentional processes, and consequently subsumed within the workings of a unified system. Attention is considered a necessary condition for learning to occur. For stimuli containing both adjacent and nonadjacent dependencies, task demands focus attention on either adjacent or nonadjacent relations, resulting in exclusive learning for the type of dependency that was *actively processed*. One potential weakness of this view is that it posits no readily accessible explanation for how learners might track different statistical information (from attending to the same elements) or be simultaneously sensitive to both adjacent and nonadjacent statistical relations within the same input stream, e.g., as indicated in Hunt and Aslin (2001) and Vuong, Meyer, and Christiansen (2011) respectively. Another difficulty is that neonates appear sensitive to statistical adjacent dependencies (in the form of sequential co-occurrences)—despite being trained / tested in the auditory modality while asleep (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Finally, Pacton and Perruchet's unified view on its own would not account for the within-subjects lack of association between adjacency and nonadjacency tasks

preliminarily reported in the small-scale individual-differences study from Chapter 2.

Instead, in chapter 3, it was proposed as a *working hypothesis* forward that individual differences in processing biases for integrating adjacent and nonadjacent dependencies might underlie the observed patterns of performances across the statistical learning and natural language tasks. Rather than necessarily entailing separate systems, such biases could be conceivably captured within the same, underlying unitary system for learning/processing. Thus, while there is sometimes a tendency for researchers to view ostensibly different behavioral manifestations as reflecting architectural dichotomies, computational frameworks (such as simple recurrent networks) suggest that functional distinctions can exist within a single system (Cleeremans, 1997). One theoretical aspect of this hypothesis is that it intrinsically accommodates a proposed continuity between acquisition and processing mechanisms in language (Chang, Dell, & Bock, 2006; Christiansen & Chater, 2008; Farmer, Christiansen, & Monaghan, 2006; Seidenberg, 1997; Seidenberg & MacDonald, 2001).¹² That is, by this view, mechanisms involved in acquiring knowledge about probabilistic dependencies (statistical learning mechanisms) are integrally shared with those involved in the online integration of probabilistic constraints during natural language processing (e.g., MacDonald, Pearlmutter,

¹² Because of the differential linking from adjacent to nonadjacent structure for the findings herein (i.e., lack of correlation between adjacent and nonadjacent learning, Chapter 2; inverse correlation between local and long-distance dependency processing, Chapter 3), one interpretative gloss might be to speculate about potential distinctions between learning mechanisms and processing mechanisms. So the theoretical compatibility of an alternate account with a continuous learning/processing framework merits mention.

& Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998).

More compellingly, across two separate groups of learners, the trajectories for adjacent statistical learning (see Figure 3.2) and nonadjacent statistical learning (see Figure 4.3) observed across our studies appear strikingly similar. Computational modeling might therefore be fruitful for better understanding the network dynamics that produce different groups of learners (i.e., behaviorally favoring adjacent statistical cues, favoring nonadjacent statistical cues, or more equi-biased towards both) from out of shared computational principles arising within the same mechanism(s). Equally vital to fleshing out such a proposal (or for developing a better framework, if the hypothesis is disconfirmed) would be closer attention to the empirical interrelationships between statistical learning and language across a variety of dependency-structures.

This initial tripartite identification of statistical learner biases also appears to accord well with a couple other findings of individual differences. Hunt and Aslin (2001), examining learning for predictive sequential relationships, concluded that there were two types of learners from their SRT task, differing “in the kinds of statistics and processing styles implemented during learning” (p. 679). Namely, learners appeared to favor either bigram or trigram information when anticipating the next element-transition in a temporal sequence. Cho, Szudlarek, Kukona, and Tabor (2011) examined parallels between human and SRN performances on predicting transitions for center-embedded recursive sequences generated by an artificial grammar. Performance variation was observed as clustering into a small set of coherent

behaviors—with some learners approximating a Simple Markov strategy (viz., adjacent dependency learning) and other learners clustering into strategies accommodating sensitivity to both adjacent and nonadjacent dependencies.

Non-language specific learning mechanisms

The empirical demonstrations herein connect individual differences in statistical learning with variations in natural language processing. As such, they implicate a substantial role for non-language specific sequence learning mechanisms in language. The general-purpose nature of such mechanisms is corroborated by the seeming ubiquity of statistical learning phenomena across myriad non-linguistic contexts, including to other perceptual, action, and (visuo)motor learning domains (Baldwin, Andersson, Saffran, & Meyer, 2008; Conway & Christiansen, 2005; Fiser & Aslin, 2002a, 2002b; Hunt & Aslin, 2001).

Additionally, it receives empirical support from other newly emerging studies on statistical learner differences. In one study, statistical learning of visual sequence regularities was positively linked to within-subject differences in adults' speech perception abilities (Conway, Bauernschmidt, Huang, & Pisoni, 2010). In two other experiments, visual statistical learning abilities were positively correlated with standardized language outcome measures for deaf children with cochlear implants (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011) and, in a larger-scale study, with syntactic priming of natural language structures in primary-school children (Kidd, 2012).

These within-subjects findings connect with former work on atypical group differences, substantiating a consistent linkage thus far between statistical learning abilities and language variation (Christiansen et al., 2010; Evans, Saffran, & Robe-Torres, 2009; Grunow, Spaulding, Gómez, & Plante, 2006; Plante, Gómez, & Gerken, 2002; Tomblin, Mainela-Arnold, & Zhang, 2007). Further, the tasks used with these groups (artificial grammar learning, sequence learning, and statistical segmentation) have been implemented in the visual modality—both within the statistical learning literature at large, as well as for a subset of the aforementioned studies [on individual and group differences]. Admittedly, the stimuli and procedural details for some of these implementations does not entirely rule out the use, in part, of auditory statistical learning via the recruitment of phonological representations or verbal encoding (see below for implications). The studies above differ in the degree to which this could be plausible (and the matter could be addressed more definitively with future experimentation). However, it seems clear that such statistical learning mechanisms are not inherently language-specific, with shared computational principles (even if realized across distinct perceptual-modality subsystems; Conway & Christiansen, 2006, 2009) and with rather basic properties (“low-level,” associative, and perceptually-grounded).

A final note: Domain-general versus perceptual-modality considerations

While the studies herein underscore an important role for non-language specific sequence learning mechanisms in accounting for language variation, the "domain-general" status of the correlations is equivocal. Thus, as noted

earlier, statistical learning phenomena have been observed for diverse learning contexts across different modalities (auditory, visual, and tactile). However, following Conway and Christiansen (2005, 2006, 2009), these effects might be mediated by independent modality-constrained statistical learning (sub)systems, rather than subserved by a singular (amodal) statistical learning system. If this view is correct, then it is unknown whether the associations between language and statistical learning observed in Chapters 2-4 would exist or remain as strong if the statistical learning tasks required the recruitment of non-phonological representations or non-speech input (e.g., a task requiring the statistical learning of abstract shapes). Moreover, Conway and Christiansen (2006) showed that individuals could learn statistical regularities from two parallel input streams along different perceptual dimensions, even within the same sense modality. This might suggest the potential for some specificity of learning-performance correlations with respect to major perceptual domains both across and *within* sensory modalities.

An intriguing future line of work, therefore, would be to probe the extent to which statistical learning in other perceptual domains relates to natural language variation, as well as to examine interrelationships among different modality-mediated statistical learning and performances on other cognitive tasks requiring the tracking of sequential dependencies. Given that statistical learning (sub)systems would appear to operate via the same computational principles (albeit with different modality biases), one possibility is that there may be a shared commonality in statistical learning skills across different perceptual-modality domains. An individual's

performances across such domains might thus be highly intercorrelated, with a possibly stronger association observed for language and statistical learning tasks implemented within the same perceptual domain (i.e., phonologically-mediated statistical learning and natural language skill [for oral languages], as well as perhaps visuomotor statistical learning and sign language skill).

Conclusion

To recap, the work presented in this thesis is among the first to establish empirical links between statistical learning and language through the framework of studying individual differences. Implicit in this hypothesized linkage is the larger perspective that language behaviors are intimately enmeshed in the operations of interacting, more basic perceptual-cognitive mechanisms that are not inherently language-specific—with statistical learning possibly being a premier demonstration of such general mechanisms. As with most newly emerging sub-areas of research, there are limitations in the conclusions that can be presently drawn. Future empirical demonstrations (e.g., larger-scale and psychometric investigations) are needed (and are currently ongoing) to substantiate this perspective. However, results from the studies herein provide initial strong evidence for an overall positive and intricate empirical relationship between statistical learning and language processing.

Secondly, the nature of reported individual differences across statistical learning and language has implications for whether the same or separate systems may be entailed in processing adjacent and nonadjacent language

dependencies. While such theorizing and experimental work is in very early stages, a reconsideration of the standard undifferentiated view of statistical learning [in Figure 1.1 *left*] might provide insights into the architectural and computational constraints underlying language and statistical learning processes. Simple recurrent network simulations, as employed in Chapter 4 and closely related work (described throughout this thesis), may provide a candidate framework for advancing our understanding.

Thirdly, statistical learning mechanisms implicate a substantial role for experience in shaping probabilistic knowledge about language (and other structured aspects of the world). Thus, results from these studies, which dovetail with other empirical work and simulations from experience-based accounts of language processing, offer a proposed synthesis between long-standing debates about the role of capacity- versus experience- based factors (and between experiential and biological factors, more generally) in language. Namely, *variations* in the *ability* to learn from *experience*—via statistical learning mechanisms—may importantly contribute to differences in language.

Finally, while work on individual differences in statistical learning is ongoing and nascent, the intricate pattern of interrelations observed thus far suggests that language and statistical learning may sometimes be related in more nuanced, counterintuitive ways than traditionally supposed. Future research that continues to empirically bridge this variation thus has the potential to open up new vistas for understanding the broad perceptual-cognitive processes upon which statistical learning and language may commonly supervene.

REFERENCES

- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382-1407.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234-272.
- Cho, P. W., Szudlarek, E., Kukona, A., & Tabor, W. (2011). An artificial grammar investigation into the mental encoding of syntactic structure. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Cognitive Science Society* (pp. 1679-1684). Austin, TX: Cognitive Science Society.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 489-509.
- Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, 116, 382-393.
- Cleeremans, A. (1997). Principles for implicit learning. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 196-234). Oxford: Oxford University Press.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114, 356-371.

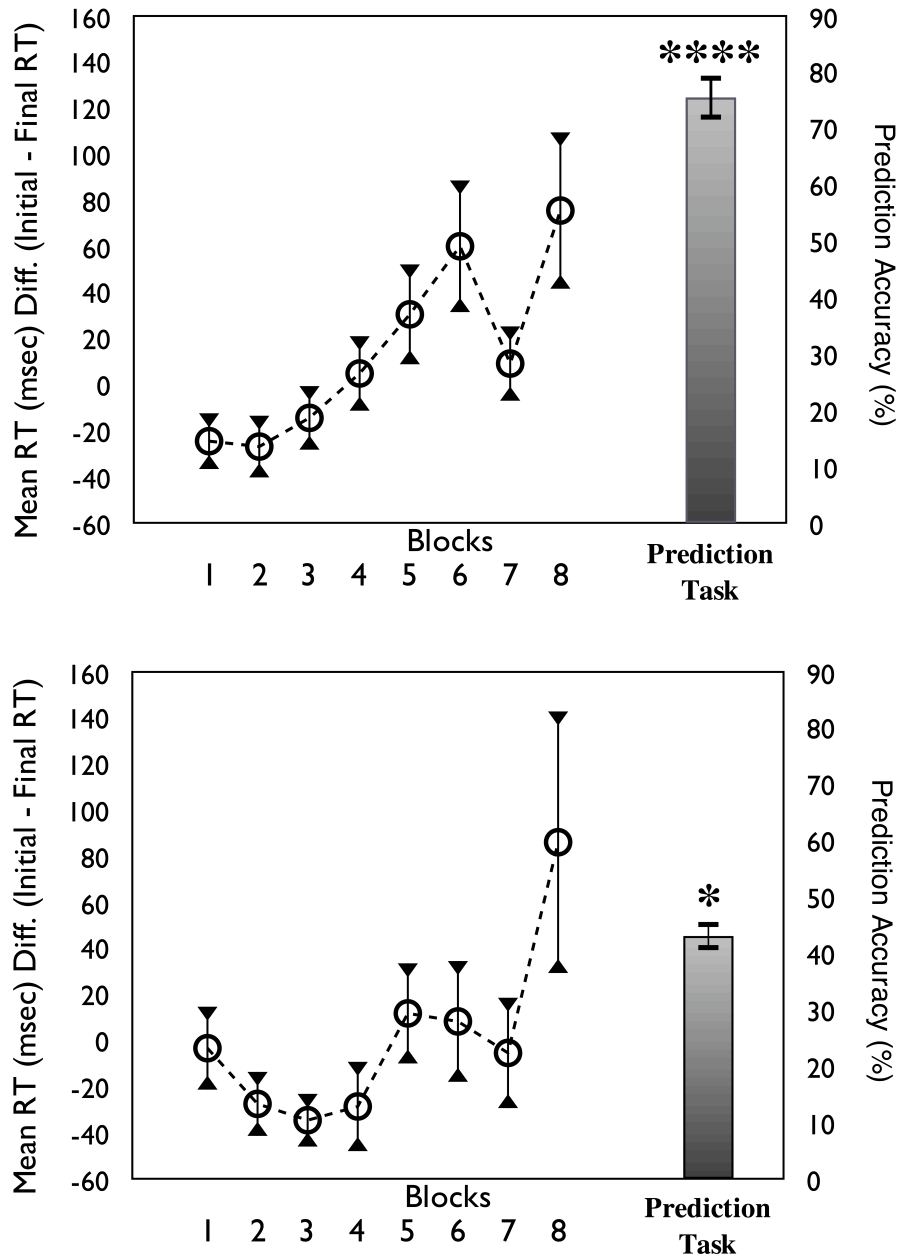
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24-39.
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, 17, 905-912.
- Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, 21, 561-580.
- Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, 14, 69-82.
- de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, 107, 763-774.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321-335.
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, 103, 12203-12208.
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458-467.

- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.
- Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, 8, 245-247.
- Friederici, A. D., Bahlmann, J., Heim, S., Schibotz, R. I., & Anwender, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, 103, 2458-2463.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39, 158-170.
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658-680.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48, 171-184.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35-54.

- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1:31. doi:10.3389/fpsyg.2010.00031
- Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 80-96.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427-456.
- Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, 28, 383-407.
- Plante, E., Gómez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35, 453-462.
- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599-1603.
- Seidenberg, M. S., & MacDonald, M. C. (2001). Constraint satisfaction in language acquisition. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 281-318). Westport, CT: Ablex.

- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10, 21.
- Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, 3, 269-293.
- Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and nonadjacent dependencies in statistical learning. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Cognitive Science Society* (pp. 964-969). Austin, TX: Cognitive Science Society.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250-271.

APPENDIX



Group learning trajectory (as a plot of mean RT difference scores) and prediction accuracy for "good" statistical learners (top panel) and "poor" statistical learners (bottom panel) from the nonadjacent statistical learning study in Chapter 4. Individuals scoring about 50% accuracy on the prediction task were classified as "good" statistical learners ($n = 17$, $M = 76\%$, $SE = 3.9$), whereas those scoring at or below 50% accuracy were classified as "poor" statistical learners ($n = 13$, $M = 43\%$, $SE = 2.5$).